

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>5</sup> : C12P 19/34, C12Q 1/68 C07H 21/04, C12M 1/38	A1	(11) International Publication Number: WO 93/08297 (43) International Publication Date: 29 April 1993 (29.04.93)
<p>(21) International Application Number: PCT/US92/09230</p> <p>(22) International Filing Date: 21 October 1992 (21.10.92)</p> <p>(30) Priority data: 781,424 23 October 1991 (23.10.91) US</p> <p>(71) Applicant: BAYLOR COLLEGE OF MEDICINE [US/US]; One Baylor Plaza, Houston, TX 77030 (US).</p> <p>(72) Inventors: LUPSKI, James, R. ; 11102 Ashcroft Drive, Houston, TX 77096 (US). KOEUTH, Thearith ; 7900 Cambridge, 4-2G, Houston, TX 77054 (US). VERSALOVIC, James ; 2600 Bellefontaine, C-25, Houston, TX 77025 (US).</p>		<p>(74) Agent: PAUL, Thomas, D.; Fulbright &amp; Jaworski, 1301 McKinney, Suite 5100, Houston, TX 77010-3095 (US).</p> <p>(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<p>(54) Title: FINGERPRINTING BACTERIAL STRAINS USING REPETITIVE DNA SEQUENCE AMPLIFICATION</p>		
<p>(57) Abstract</p> <p>Oligonucleotide primers and methods for indentifying strains of bacteria by genomic fingerprinting are described. The methods are applicable to a variety of samples. The testing procedure includes amplifying the bacterial DNA in the sample to be tested by adding a pair of outwardly-directed primers to the sample. The primers are capable of hybridizing to repetitive DNA sequences in the bacterial DNA and extending outwardly from one hybridizable repetitive sequence to another hybridizable repetitive sequence. After amplification the extension products are separated by size and the specific strain of bacteria is determined by measuring the pattern of sized extension products. The procedure to identify strains of bacteria by fingerprinting has a variety of uses including: identifying bacteria in infections, agriculture and horticulture plots, bioremediation, food monitoring, production monitoring, quality assurance and quality control.</p>		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SK	Slovak Republic
CI	Côte d'Ivoire	LK	Sri Lanka	SN	Senegal
CM	Cameroon	LU	Luxembourg	SU	Soviet Union
CS	Czechoslovakia	MC	Monaco	TD	Chad
CZ	Czech Republic	MG	Madagascar	TG	Togo
DE	Germany	ML	Mali	UA	Ukraine
DK	Denmark	MN	Mongolia	US	United States of America
ES	Spain			VN	Viet Nam
FI	Finland				

-1-

## FINGERPRINTING BACTERIAL STRAINS USING REPETITIVE DNA SEQUENCE AMPLIFICATION

### FIELD OF THE INVENTION

5           The present invention relates generally to the use of oligonucleotide probes directed to repetitive DNA sequence elements to identify bacteria. More particularly, it relates to the use of these probes as primers for the amplification of bacterial genomic DNA between repetitive sequences, and the use of these amplification products to construct DNA fingerprints  
10           unique to the probed genome. It also relates to the disclosure of specific primers which are useful as oligonucleotide probes in the practice of this invention.

### BACKGROUND OF THE INVENTION

15           Interspersed repetitive DNA sequence elements have been characterized extensively in eucaryotes although their function still remains largely unknown. The conserved nature and interspersed distribution of these repetitive sequences have been exploited to amplify unique sequences between repetitive sequences by the polymerase chain reaction. Additionally, species-specific repetitive DNA elements have been  
20           used to differentiate between closely related murine species.

          Prokaryotic genomes are much smaller than the genomes of mammalian species (approximately  $10^6$  versus  $10^9$  base pairs of DNA, respectively). Since these smaller prokaryotic genomes are maintained through selective pressures for rapid DNA replication and cell  
25           reproduction the non-coding repetitive DNA should be kept to a minimum unless maintained by other selective forces. For the most part prokaryotes have a high density of transcribed sequences. Nevertheless, families of short intergenic repeated sequences occur in bacteria.

-2-

The presence of repetitive sequences has been demonstrated in many different bacterial species. Reports of novel repeated sequences in the eubacterial genera, *Escherichia*, *Salmonella*, *Deinococcus*, *Calothrix*, and *Neisseria*, and the fungi, *Candida albicans* and *Pneumocystis carinii*, illustrate the presence of dispersed extragenic repetitive sequences in many organisms. One such family of repetitive DNA sequences in eubacteria is the Repetitive Extragenic Palindromic (REP) elements. The consensus REP sequence for this family includes a 38 mer sequence containing six totally degenerate positions, including a 5 bp variable loop between each side of the conserved stem of the palindrome. Another family of repetitive elements is the Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences. ERIC is larger (consensus sequence is 126-mer) and contains a highly conserved central inverted repeat. The ERIC and REP consensus sequences do not appear to be related.

Previous studies have used repeated rRNA genes as probes in Southern blots to detect restriction fragment length polymorphisms (RFLPs) between strains. Repeated tRNA genes have been used as consensus primer binding sites to directly amplify DNA fragments of different sizes by PCR amplification of different strains. Limitations of both techniques include the use of radioisotope and time-intensive methods such as Southern blotting and polyacrylamide gel electrophoresis to clearly distinguish subtle differences in the sizes of the DNA fragments generated. The latter technique could only distinguish organisms at the species and genus level. The tDNA-PCR fingerprints are generally invariant between strains of a given species and between related species. Other previous studies include the use of species-specific repetitive DNA elements as primer-binding sites for PCR-based bacterial species identification. Though such methods allow species identification by PCR with picogram amounts of DNA, only single PCR products are generated which precludes the generation of strain-specific genomic fingerprints.

-3-

Although these previous studies demonstrated that species-specific repetitive DNA elements can be used as primer-binding sites for PCR-based bacterial species identification, these methods only generated single PCR products in a single species. The present invention provides a novel approach to using extragenic repetitive sequences to directly fingerprint bacterial genomes. Analysis of amplification products resulting from amplifying unique sequences between primers to bacterial DNA repeat sequences, reveals unique distances between repeat sequences. This pattern of distances uniquely fingerprints different bacterial species and strains. Thus, this approach provides a quick and reliable method to type bacteria by genomic fingerprinting.

#### SUMMARY OF THE INVENTION

An object of the present invention is a method of identifying a strain of bacteria by amplifying the DNA between repetitive DNA sequences and measuring the pattern of sized extension products.

An additional object of the present invention is provision of primer pairs to bacterial repetitive DNA sequences.

A further object of the present invention is a method of identifying a strain of bacteria in samples from physiological and non-physiological sources.

An additional object of the present invention is a method for diagnosing bacterial disease in humans and animals.

A further object of the present invention is the detection of bacterial disease or contamination in plants.

An additional object of the present invention is the monitoring of bacterial contamination in foods.

A further object of the present invention is a method for developing a library of fingerprints to identify specific strains of bacteria.

An additional object of the present invention is a method for monitoring bacterial contamination in soils, liquids, solids and other samples from environmental sources.

5 A further object of the present invention is a method for monitoring manufacturing processes for bacterial contamination.

An additional object of the present invention is a method for quality assurance or quality control of microbiological based laboratory assays.

A further object of the present invention is a method for genomic mapping.

10 An additional object of the present invention is the monitoring of bacterial populations in bioremediation sites.

A further object of the present invention is the monitoring of bacterial infections.

15 An additional object of the present invention is a method for the automated identification of a bacterial strain.

A further object of the present invention is a machine for the automated identification of bacterial strains.

20 Thus, in accomplishing the foregoing objects, there is provided in accordance with one aspect of the present invention a method for identifying a strain of bacteria in a sample, comprising the steps of: amplifying DNA between repetitive sequences in the bacteria by adding a pair of outwardly-directed primers to the sample, which primers are capable of hybridizing to repetitive DNA sequences in the bacterial DNA, and extending outwardly from one hybridizable repetitive sequence to  
25 another hybridizable repetitive sequence; separating the extension products generated in the amplification step by size; and determining the specific strain of bacteria by measuring the pattern of sized extension products.

30 In specific embodiments of the present invention the primers are between about 10 to 29 mer and preferably between about 15 to 25 mer.

The primers can be specific to any repetitive sequence but in the preferred embodiments are specific to ERIC, REP, Ngrep or Drrep.

In various aspects of the present invention the method can be used for: (1) diagnosis of bacterial disease, in plants animals and humans; (2) monitoring for bacterial content and/or contamination in the environment; (3) monitoring food for bacterial contamination; (4) monitoring manufacturing processes for bacterial contamination; (5) monitoring quality assurance/quality control of laboratory tests involving microbiological assays; (6) tracing bacterial contamination and/or outbreaks of bacterial infections; (7) genome mapping; (8) monitoring bioremediation sites; and (9) monitoring agricultural sites for test crops, bacteria and recombinant molecules.

The method is useful on pure or isolated cultures as well as actual samples from the test site. In a preferred embodiment multiple primers to different repetitive DNA can be used.

Because of the simplicity of the test it can also be automated for rapid and quick assay of samples.

A further aspect of the present invention is a machine for automating the identification of bacterial strains.

Other and further objects, features and advantages will be apparent from the following description of the presently preferred embodiments of the invention, which are given for the purpose of disclosure, when taken in conjunction with the accompanying drawings.

#### DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic showing the binding of outwardly-directed primers.

Figure 2 shows the alignment of various REP oligonucleotide primer sequences with respect to a REP consensus sequence.

Figure 3 shows the alignment of ERIC oligonucleotide primer sequences with respect to the central inverted repeat of an ERIC consensus sequence.

5

Figure 4 shows PCR amplification of *E. coli* strain W3110 genomic DNA with different REP and ERIC oligonucleotide primer sets.

Figure 5 is a 1% agarose gel demonstrating the specificity of ERIC oligonucleotide primer/template interactions.

Figure 6 shows the results of REP-PCR of strains within the Gram-negative enterobacterial species.

10

Figure 7 shows the results of ERIC-PCR of strains within the Gram-negative enterobacterial species.

Figure 8 shows a "bug blot" hybridization of REP in a wide variety of bacteria.

Figure 9 shows the evolutionary conservation of REP sequences.

15

Figure 10 shows a "bug blot" hybridization of ERIC in a wide variety of bacteria.

Figure 11 shows the evolutionary conservation of ERIC sequences.

Figure 12 shows the evolutionary conservation of Ngrep.

Figure 13 shows REP/ERIC Fingerprinting of *B. subtilis*.



-7-

Figure 14 is REP-PCR of *E. coli* W3110 genomic cosmid library.

Figure 15 is REP-PCR of *E. coli* W3110 genomic cosmid library.

Drawings are not necessarily to scale. Certain features of the invention may be exaggerated in scale or shown in schematic form in the interest of clarity and conciseness.

#### DETAILED DESCRIPTION OF THE INVENTION

It will be readily apparent to one skilled in the art that various substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

DNA amplification as used herein refers to any process which increases the number of copies of a specific DNA sequence. A variety of processes are known. One of the most commonly used is the Polymerase Chain Reaction (PCR) process of Mullis as described in U.S. Patent Nos. 4,683,195 and 4,683,202 both issued on July 28, 1987. In general the PCR amplification process involves an enzymatic chain reaction for preparing exponential quantities of a specific nucleic acid sequence. It requires a small amount of a sequence to initiate the chain reaction and oligonucleotide primers which will hybridize to the sequence. In PCR the primers are annealed to denatured nucleic acid followed by extension with an inducing agent (enzyme) and nucleotides. This results in newly synthesized extension products. Since these newly synthesized sequences become templates for the primers, repeated cycles of denaturing, primer annealing, and extension results in exponential accumulation of the specific sequence being amplified. The extension product of the chain reaction will be a discrete nucleic acid duplex with a termini corresponding to the ends of the specific primers employed. In the present invention the extension product traverses from one repetitive sequence to another

repetitive sequence. Since the repetitive sequences are interspersed throughout the genome at different distances from each other, there will be exponential growth of all the different sizes. The pattern of extension products of different sizes provides a specific fingerprint for each bacteria.

5           The term "oligonucleotide primer" as used herein defines a molecule comprised of more than three deoxyribonucleotides or oligonucleotides. Its exact length will depend on many factors relating to the ultimate function and use of the oligonucleotide primer, including temperature, source of the primer and use of the method. The oligonucleotide primer  
10       can occur naturally (as a purified fragment or restriction digestion product) or be produced synthetically. The oligonucleotide primer is capable of acting as an initiation point for synthesis, when placed under conditions which induce synthesis of a primer extension product complementary to a nucleic acid strand. The conditions can include the  
15       presence of nucleotides and an inducing agent such as a DNA polymerase at a suitable temperature and pH. In the preferred embodiment the primer is a single-stranded oligodeoxyribonucleotide of sufficient length to prime the synthesis of an extension product from a specific sequence in the presence of an inducing agent. In the present application in the  
20       preferred embodiment the oligonucleotides are usually between about 10 mer and 29 mer. In the preferred embodiment they are between 15 and 25 mer. Sensitivity and specificity of the oligonucleotide primers are determined by the primer length and uniqueness of sequence within a given sample of a template DNA. Primers which are too short, for  
25       example, less than 10 mer may show non-specific binding to a wide variety of sequences in the genomic DNA and thus are not very helpful.

Each primer pair herein is selected to be substantially complementary to the different strands of each specific repetitive sequence to which the primer pairs bind. Thus one primer of each pair is  
30       sufficiently complementary to hybridize with a part of the sequence in the

sense strand and the other primer of each pair is sufficiently complementary to hybridize with a different part of the same repetitive sequence in the anti-sense strand.

5 It should also be recognized that a single primer can be considered a primer pair in this invention. Because the primer binds to repetitive sequences and because the repetitive sequences can be orientated in both directions, a single primer can bind to both strands of a repetitive sequence and amplify the sequence between two separate repetitive sequences.

10 As used herein the term "outwardly directed" primer pair refers to the oligonucleotide primers and their binding as seen in Figure 1. In the present application one primer is substantially complementary to the sense strand. This primer binds to the sense strand in such an orientation that the extension product generated from the 3' end of the primer extends  
15 away from the repetitive DNA sequence to which the oligonucleotide primer is bound and across the non-repetitive DNA to a second repetitive DNA sequence. The other member of the primer pair binds to the anti-sense strand. This primer binds in an orientation such that extension products generated on the 3' end extends away from the repetitive DNA  
20 sequence to which the primer is bound and across the non-repetitive DNA to the next repetitive DNA sequence. Thus, within a specific repetitive DNA sequence the primer pair is bound to the complementary DNA strands 5' to 5' (see Figure 1) and, thus, the extension products grow away from each other across the non-repetitive DNA. The extension  
25 products from the two paired primers are complementary to each other and can serve as templates for further synthesis by binding the other member of the primer pair.

As used herein the term "extension product" refers to the nucleotide sequence which is synthesized in the presence of nucleotides and an  
30 inducing agent such as a polymerase from the 3' end of the

-10-

oligonucleotide primer and which is complementary to the strand to which the oligonucleotide primer is bound.

As used herein the term "differentially labelled" shall indicate that the extension product can be distinguished from all the others because it has a different label attached or is of a different size or binds to a specific oligonucleotide or a combination thereof. One skilled in the art will recognize that a variety of labels are available. For example, these can include radioisotopes, fluorescers, chemluminescers, enzymes and antibodies. Various factors affect the choice of the label. These include the effect of the label on the rate of hybridization and binding of the primer to the DNA, the sensitivity of the label, the ease in making the labeled primer, probe or extension products, the ability to automate, available instrumentation, convenience and the like. For example in one embodiment of the present invention size alone is sufficient to distinguish the patterns and thus no other label is needed. The size differences can be determined after staining the DNA, for example with ethidium bromide. However, when detecting multiple species in a sample or for multiple repetitive sequences it may be advantageous to add a radioactive label such as  $^{32}\text{P}$ ,  $^3\text{H}$  or  $^{14}\text{C}$ ; a different fluorescer such as fluorescein, tetramethylrhodamine, Texas Red or 4-chloro-7-nitrobenzo-2-oxa-1-diazole (NBD); or a mixture of different labels such as radioisotopes, fluorescers and chemluminescers.

The term "repetitive DNA" as used herein refers to non-coding sequences of DNA containing short repeated sequences and dispersed throughout the bacterial genome.

As used herein (1) "REP" refers to the repetitive extragenic palindromic elements. The REP consensus sequence is shown in SEQ. ID. NO. 1. (2) "ERIC" refers to the enterobacterial repetitive intergenic consensus sequence. The ERIC consensus sequence is shown in SEQ. ID. NO. 36. (3) "Ngrep" refers to the *Neisseria* repetitive elements. The

-11-

Ngrep consensus sequence is shown in SEQ. ID. NO. 46. (4) "Drrep" refers to the *Deinococcus* repetitive elements. The Drrep consensus sequence is shown in SEQ. ID. NO. 60. These repetitive elements are found interspersed throughout the bacterial genome. These four sequences or any combination of these four sequences can be used in the present invention. Further, one skilled in the art will understand that as new repetitive sequences in bacteria become known they can also be used in the method of the present invention. By binding outwardly-directed primers to these repetitive sequences and performing amplification one can generate unique fingerprints and identify individual strains of bacteria.

The oligonucleotide primers may be prepared using any suitable method known in the art. For example the phosphodiester, and phosphotriester methods or automated embodiments thereof. It is also possible to use a primer which has been isolated from biological sources such as with a restriction endonuclease digest.

The repetitive sequence to which the primers bind can be selected from any of the repetitive regions that are found in bacteria. The repetitive sequences can be identified by a variety of methods. This may be done manually by comparing the sequences of the published nucleic acid sequences for bacterial genomes. A more convenient method, however, is to use a computer program to compare the sequences. In this way one can generate a consensus DNA sequence for use in the methods of the present application.

Any source of bacterial nucleic acid in purified or non-purified form can be utilized as starting material, provided it contains or is suspected of containing a bacterial genome of interest. Thus, the bacterial nucleic acids may be obtained from any source which can be contaminated by bacteria. When looking for bacterial infection or in distinguishing bacteria from human or animal subjects, the sample to be tested can be selected or

-12-

extracted from any bodily sample such as blood, urine, spinal fluid, tissue, vaginal swab, stool, amniotic fluid or buccal mouthwash.

In other applications the sample can come from a variety of sources. For example: (1) in horticulture and agricultural testing the sample can be a plant, fertilizer, soil, liquid or other horticultural or agricultural product; (2) in food testing the sample can be fresh food or processed food (for example infant formula, seafood, fresh produce and packaged food); (3) in environmental testing the sample can be liquid, soil, sewage treatment, sludge and any other sample in the environment which is considered or suspected of being contaminated by bacteria.

When the sample is a mixture of material for example blood, soil and sludge it can be treated within an appropriate reagent which is effective to open the cells and expose or separate the strands of nucleic acids. Although not necessary, this lysing and nucleic acid denaturing step will allow amplification to occur more readily. Further, if desired, the bacteria can be cultured prior to analysis and thus a pure sample obtained. This is not necessary, however.

The inducing agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products. Examples of inducing enzymes which have been used for this purpose include *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase I, T4 DNA polymerase, *Taq* DNA polymerase, Vent DNA polymerase and other available DNA polymerases.

As used herein "fingerprinting" refers to the fact that each strain of bacteria has its characteristic size pattern of extension products which can be used to identify the bacterial strain. This unique pattern is each strain's genomic fingerprint.

One embodiment of the present invention includes a method for identifying a strain of bacteria in a sample, comprising the steps of: amplifying DNA by adding a pair of outwardly-directed primers to the

-13-

sample, wherein the primers are capable of hybridizing to repetitive DNA sequences in the bacterial DNA and extending outwardly from one hybridizable repetitive sequence to another hybridizable repetitive sequence; separating the extension products generated in the amplification step by size; and determining the specific strain of bacteria by measuring the pattern of sized extension products.

It will be recognized that the separating step of this embodiment may be accomplished by any number of techniques and methods which will separate the extension products by size. Examples include but are not limited to gel electrophoresis, capillary electrophoresis, chromatography, pulsed field gel electrophoresis and mass spectrometry. Thus, one skilled in the art will recognize that the separation of extension products can be done by a variety of methods. The choice of method will depend on a number of factors, such as the available laboratory equipment, the amount of extension product present, the label if any, the dye, the preference of the party performing the testing, convenience and the like.

Capillary electrophoresis allows the rapid separation of DNA fragments through tiny polyacrylamide gels in thin capillaries. The chief advantage is that much larger voltages can be applied and resolution is enhanced. The process can be automated. Once tubes are loaded, electrophoresis and data acquisition can be automated by direct connection to computer. An example includes the Model 270A-HT High Throughput Capillary Electrophoresis System (Applied Biosystems). Instead of bands on a gel, the DNA fragments are represented by spikes as a function of time indicating the presence of different molecules of different sizes. Another advantage is that not only can PCR-generated spike patterns be quickly obtained with greater resolution of different-sized fragments, but intensity of different bands could be accurately quantitated; permitting even greater resolution.

-14-

Non-electrophoresis methods, namely chromatography, can be used to separate PCR-generated DNA fragments by size. High-Performance Liquid Chromatography (HPLC) methods can be used to separate DNA fragments by the use of size-exclusion columns (Series 800 HRLC Gradient System - BioRad). DNA fragments are represented by spikes as a function of time and the data is digitized and fed directly to a computer. Electrophoresis methods, however, are usually preferred because of greater reliability and resolution.

One skilled in the art will recognize that measurement of the pattern of sized extension products to determine the specific strain of bacteria present may also be accomplished by several means, direct visualization or by automation using a bar code reader, a laser reader, digitizer, a photometer, a fluorescence reader or computer planimetry. The choice of measurement method depends in part on the separation step and available instrumentation.

A variety of primers can be used to detect repetitive sequences in bacteria. The primers depend on which repetitive sequence is being detected and which bacteria are being detected. One skilled in the art can readily determine which repetitive sequence and primers to use depending on the bacteria being examined. In the embodiment of the present invention primer pairs have been selected from the sequences in the groups consisting of SEQ. ID. NOS. 4 to 35, 38 to 45 and 48 to 57. In the preferred embodiment when REP is being used for the primer annealing site one member of the pair of primers is selected from the group consisting of SEQ. ID. NOS. 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 34 and the other member of the pair is selected from the group consisting of SEQ. ID. NO. 7, 8, 9, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, and 35. In the most preferred embodiment SEQ. ID. NO. 4 and SEQ. ID. NO. 7 are used.



-15-

When the repetitive sequence is ERIC one member of the pair of primers is selected from this group consisting of SEQ. ID. NO. 38, 39, 40, and 41 and the other member of the pair is selected from SEQ. ID. NOS. 42, 43, 44 and 45 and in the most preferred embodiment SEQ. ID. NO. 38 and SEQ. ID. NO. 42 are used.

When the repetitive sequence is Ngrep one member of the pair of primers is selected from the group consisting of SEQ. ID. NOS. 48, 49, 50 and 51 and the other member of the pair is selected from the group consisting of SEQ. ID. NOS. 52, 53, 54 and 55. In the preferred embodiment SEQ. ID. NO. 48 and SEQ. ID. NO. 52 are used.

When the repetitive sequence is Drrep only a single primer is used. The primer is either SEQ. ID. NOS. 56 or 57. This is an example where a single primer acts as a pair. Use of both primers will not result in unique fingerprint patterns.

One skilled in the art will readily recognize that as more repetitive sequences are determined the primer pair which gives the best fingerprint pattern can be easily selected. For example, a primer to the new sequence is synthesized and the method of the present invention is seen. After examining the resulting pattern from each primer pair the primer pair which best distinguishes the specific test strains can be identified.

In addition to the above described method a plurality of pairs of primers can be added to the method. Each pair of primers will bind a different repetitive sequence. For example, any combination of two or more of each of REP, ERIC, Ngrep and Drrep primer pairs can be added. Further, the multiprimer assay can be enhanced by differentially labeling the primer pairs. Thus, after amplification, not only can the sized pattern be examined, but the size pattern for each label can be examined. For example, REP and ERIC oligonucleotide primers can be used. Each is labeled with a different fluorescent label. The resultant differential

-16-

labeling pattern can be determined by fluorescence scanning. This procedure can provide finer fingerprint patterns.

5 After electrophoresis in polyacrylamide gels, the gels are scanned by a laser-based fluorescence detector and the results digitized directly by computer connected to the detector. Further, using a Genescanner (Applied Biosystems) allows the entire process to be automated.

10 One skilled in the art will readily recognize that this method has many advantages. It can be readily modified for automated identification of strains of bacteria. In one embodiment the amplification is done in an auto-PCR instrument, the extension products are removed and separated on a sizing gel or by chromatography. The sizing pattern is determined by an automatic reader and each pattern can be recognized by a computer means. The computer will store fingerprints of known bacteria for comparison with test results. In the automated method, bar code readers, 15 laser readers, digitizers, photometers, fluorescent readers and computer planimetry can be used to help automate the system.

In another embodiment of the present invention there is a kit for determining the identity of strains of bacteria. This kit comprises a container having a pair of outwardly-directed PCR primers to a repetitive 20 sequence in bacteria. This kit can have any of the PCR primers selected from the group consisting of SEQ. ID. NO. 4 to 35, 38 to 45 and 48 to 57 or combination thereof. One skilled in the art will readily recognize that the number and type of primers which are in the kit will depend on the use of the kit as well as the sequences which are to be detected.

25 A further embodiment of the present invention is a machine for identifying a strain of bacteria comprising an automated PCR amplifying means, a separation means, a sampling means for removing the extension products from the PCR means and transferring them to the separation means, a reading means for measuring patterns of extension products 30 after separation of the separation means, a computer means for recording

-17-

the results of the reading means and for outputting the pattern of and identifying the strain of bacteria.

5 A number of automated PCR amplifying means are known on the market. For instance a thermal cycler can be used. There are a number of arms or robotic devices and other automatic pipette and sampling machines which can be used as a sampling means for removing the extension products from the PCR reaction at the appropriate times and transferring the sample for either chromatography, gel or capillary electrophoresis, mass spectrometry or other methods or techniques used to separate the samples. In the preferred embodiment the separator means is regulated by the computer. After the separation the reader means is used to measure the pattern. The reader means will depend on the type of separation which is being used. For instance a wavelength densitometer reader or a fluorescence reader can be used depending on the label being detected. A radioisotope detector can be used for radioisotope labeled primers. In mass spectrometry the ions are detected in the spectrometer. A gel can be stained and read with a densitometer. The computer regulates the automated PCR amplification procedure, the sampling and removal from PCR, the automatic separation and reading of the samples and can be used to interpret the results and output the data.

15 The methods, instruments and procedures described herein can be used for a variety of purposes. Because of the sensitivity and specificity of the test one skilled in the art will readily recognize uses for this methodology. What follows is not an inclusive list of uses but only a sampling of specific areas where a current need exists for a quick and reliable test.

25 The methodology of the present invention can be used for diagnosing bacterial diseases whether it is in plants, animals or humans. Not only can the disease be diagnosed but the specific strain involved can be identified.

30

-18-

The environment can be monitored for bacterial contamination. The procedure works on variety of samples including liquids, sludge, sewage treatment plant samples and soil. Thus, anywhere that there has been environmental contamination that needs monitoring, the test will work.

This procedure should be very useful in the area of monitoring food contamination. A variety of producers of foods currently test their products for bacterial contamination. This procedure will help facilitate this testing. For example infant formula, seafood, fresh produce and processed food can all be readily tested by this procedure. This procedure can also be useful to detect the source of food poisoning.

Another important use of this method is in the monitoring of the bacterial populations in a bioremediation site. Bioremediation usually uses specific bacterial populations to destroy the contamination. The bacteria can be from the natural population growing at the site or bacteria added to the site to enhance the breakdown. The bacteria used in the enhancement procedures usually are from cultures and/or sludge. In any of these instances it is important to monitor the population of bacteria in the bioremediation state to make sure that the appropriate strain(s) of bacteria is present and growing. This procedure allows the rapid and quick identification of the bacteria in the population so that it can be readily monitored. The test works on samples of soil, liquid, sludge or other material to be added to the bioremediation site.

In the areas of horticulture and agriculture a variety of uses of this method are found. One can monitor bacterial inoculations of plants or bacterial disease of plants. It can also be used to monitor the distribution of recombinant bacteria added to the environment. Samples can come from the soil where bacteria have been added, or from fertilizer to make sure that the fertilizer has the appropriate bacteria. It can be used to monitor pest control where bacteria are added in order to kill pests such

as insects. This procedure allows quick and accurate monitoring of the application of the bacterial insecticide and the activity of the insecticide. Thus, in any horticulture or agriculture procedure which requires the addition of bacteria the bacteria can be monitored throughout the procedure.

Another application of this method is in the manufacturing process. A number of manufacturing processes for instance drugs, microorganism-aided synthesis, food manufacturing, chemical manufacturing and fermentation process all rely either on the presence or absence of bacteria. In either case the method of the present invention can be used. It can monitor bacterial contamination or test that strain purity is being maintained.

This method can also be used to test stored blood for bacterial contamination. This would be important in blood banking where bacteria such as *Yersinia enterocolitica* can cause serious infection and death if it is in transfused blood.

The procedure can also be used for quality assurance and quality control in monitoring bacterial contamination in laboratory tests. For example the Guthrie bacterial inhibition assay uses a specific strain of bacteria to measure phenylalanine in newborn screening. If this strain changes it could affect test results and thus affect the accuracy of the newborn screening program. This method of the present invention can be used to monitor the strain's purity. Any other laboratory test which uses or relies on bacteria in the assay can be monitored. The laboratory or test environment can also be monitored for bacterial contamination by sampling the lab and testing for specific strains of bacteria.

This procedure will also be useful in hospitals for tracing the origin and distribution of bacterial infections. It can show whether or not the infection of the patient is a hospital-specific strain. The type of treatment

-20-

and specific anti-bacterial agent can depend on the source and nature of the bacteria.

The following examples are offered by way of example, and are not intended to limit the scope of the invention in any manner.

5

### Example 1

#### Isolation and Quantitation of Genomic DNA

##### 1. Genomic DNA from Gram-negative and spirochete bacteria.

Cells were pelleted and washed twice in 1 ml of 1M NaCl by centrifugation in a fixed angle microfuge at 15,000 rpm for 5 min. Cells were washed twice and resuspended in TE (10mM Tris, 25mM EDTA, pH 8.0) and incubated in 0.2 mg/ml lysozyme and 0.3 mg/ml RNase A for 20 min at 37°C. If lysis by lysozyme was not visible with refractory pathogenic strains, 0.6% SDS was added. To these suspensions, 1% Sarkosyl and 0.6 mg/ml proteinase K were added, and the cells were incubated for 1 hr at 37°C. Cell lysates were extracted twice with phenol and twice with chloroform. The aqueous phase was precipitated with 0.33M NH<sub>4</sub> acetate and 2.5 volumes of ethanol. Precipitated threads of DNA were removed with a sterile Pasteur pipette tip, and dissolved in TE (10mM Tris, 1 mM EDTA, pH 8.0).

20

##### 2. Genomic DNA from Gram-positive bacteria.

Concentrated cell pellets were washed twice in 1 M NaCl and twice in TE (50 mM Tris, 50 mM EDTA, pH 7.8) and spun in a fixed-angle microfuge for 5 min. Cell pellets were resuspended in TE and incubated with 250 U/ml mutanolysin and 0.3 mg/ml RNase A for 30 min at 37°C. To this reaction, 0.6% SDS and 0.6 mg/ml proteinase K were added, and the mixture was incubated for 1 hr at 37°C, followed by 65°C for 45 min. Lysates were extracted twice with phenol and twice with chloroform.

25

-21-

Chromosomal DNA was precipitated and dissolved as described for Gram-negative bacteria.

5 In both instances the genomic DNA was quantitated by spectrofluorimetry at excitation and emission wavelengths of 365 nm and 460 nm respectively using the DNA-specific dye, Hoechst 33258 and a fluorometer.

### Example 2

#### Primer Synthesis

10 The oligonucleotide primers were synthesized by the phosphoramidite method using an automated DNA synthesizer, and DNA sequence information from consensus sequence data. The primers were labeled by 5' end-labeling of each oligonucleotide as described by Maniatis et al. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1982). Fifty pmol of each  
15 primer were used with 20 U T4 polynucleotide kinase and 5  $\mu$ l of  $\gamma$ -<sup>32</sup>P-ATP (6000 Ci/mmol). Labeled DNA was separated from unincorporated isotope by diluting the 50  $\mu$ l reaction volume to 1.0 ml in millipore water, followed by centrifugation of this solution through Centricon-3 (Amicon) tubes. The retained solution contains the hybridization probe.  
20 Oligonucleotides were quantitated by UV-VIS spectrophotometry with absorption measured at 260 nm.

### Example 3

#### Primer Design

1. REP oligonucleotide sequences are shown in SEQ. ID. NOS. 1 to 35 and Fig. 2. Degenerate 38-mer REPALL (SEQ. ID. NOS. 2 and 3)  
25 probes were designed which encompassed the entire consensus REP sequence (SEQ. ID. NO. 1). Other REP oligonucleotide probe pairs, each representing part of the conserved consensus sequence, were designed

-22-

with opposite orientations such that the 3' ends were directed outwards from each REP sequence. This design constituted a pair of outwardly-directed primers. If the primer to one side of the sequence is shorter than the other, inosines may be added to make the lengths of the pair equal.

5 Total degeneracy is represented either by any one of the four common bases (A, G, C, or T) placed at specific positions, or inosines placed at specific positions. Inosine contains the purine base, hypoxanthine, and is capable of forming Watson-Crick base pairs with A, G, C, or T. Positions can be partially degenerate with two of these four bases placed at specific

10 positions as chosen from the consensus REP sequence.

2. ERIC oligonucleotide sequences are shown in SEQ. ID. NOS. 36 to 45 and Fig. 3. The ERICALL oligonucleotide is a 44 mer of SEQ. ID. NO. 37 from position 42 to 85 and contains the entire conserved central core inverted repeat. Non-degenerate, consensus ERIC1R (SEQ.

15 ID. NO. 38) and ERIC2 (SEQ. ID. NO. 42) oligonucleotides were designed from each half of this core inverted repeat with opposite orientations such that the 3' ends are directed outwards from the center of the ERIC element.

3. Ngrep sequences are shown in SEQ. ID. NOS. 46 to 55.

20 SEQ. ID. NO. 46 is the consensus sequence.

4. Drrep sequences are shown in SEQ. ID. NOS. 56, 57 and 60. SEQ. ID. NO. 60 is the consensus sequence.

#### Example 4

##### Membrane Supported Hybridization

25 A single membrane containing genomic DNA from 39 different eubacterial species representing 7 of 10 different phyla of eubacteria as defined by Woese, *Microbiol Rev.*, 51:221-271 (1987), based on rDNA sequence comparisons, was named the "bug blot." This bug blot was made by adding 100 ng of denatured genomic DNA per slot, from each species



-23-

listed in Figs. 8 or 10, on Gene Screen Plus (Du Pont) membranes. The bug blot represents a slot blot DNA:DNA hybridization of genomic DNA probed with  $^{32}\text{P}$ -end labeled SEQ. ID. NO. 3 (Fig. 8) or SEQ. ID. NO. 42 (Fig. 10). These membranes were pretreated as described in Maniatis et al. Genomic DNAs were denatured in solution by heating to  $100^\circ\text{C}$  for 5 min. DNA samples were then applied to the membrane, and  $500\ \mu\text{l}$   $0.4\text{N}$  NaOH were added to each slot. Membranes were rinsed in  $1\text{X}$  SSC, and blotted dry with Whatman paper. Membranes were baked at  $80^\circ\text{C}$ . for 1 hr, and stored in sealed plastic bags at  $-20^\circ\text{C}$ .

The hybridization solution was prepared as described in Noda, A, et al. *Biotechniques* 10:474-477 (1991) for use with oligonucleotide probes on a membrane containing ordered lambda phages representing the *E. coli* W3110 genome. For REP oligonucleotide hybridization, membranes were prehybridized at  $42^\circ\text{C}$  for 1.5 hrs. The probe was denatured at  $100^\circ\text{C}$  for 5 min. Probe was added at  $1 \times 10^6\text{cpm/ml}$  hybridization solution and the membranes were incubated at  $42^\circ\text{C}$  for 15 hrs. ERIC oligonucleotide prehybridizations and hybridizations were both performed at  $65^\circ\text{C}$ . After incubation the membranes were washed twice at room temperature for 10 min with  $2\text{X}$  SSPE and  $0.1\%$  SDS, followed by one final wash (REP,  $37^\circ\text{C}$  for 15 min; ERIC,  $40^\circ$  for 1 min). Autoradiograms were exposed on Kodak X-Omat film with two intensifying screens at  $-85^\circ\text{C}$  for 24 hrs.

#### Example 5

##### DNA Amplification

There are a number of DNA amplification methods available in the art. These generally depend on the use of one or more of a variety of polymerases to catalyze DNA chain extension (polymerization) from component nucleotide bases. An example of a DNA amplification technique is the polymerase chain reaction (PCR) used here to catalyze

the extension of the oligonucleotide primer from the 3' end along the DNA template to which the primer is hybridized.

Each 25  $\mu$ l of PCR reaction contained 50 pmol each of 2 opposing primers, 100 ng of template (genomic) DNA, 1.25 mM of each of 4 dNTPs, 2 U AmpliTaq DNA polymerase (Perkin-Elmer/Cetus) in a buffer with 10% DMSO (v/v). PCR amplifications were performed in an automated thermal cycler, with an initial denaturation at 95°C for 7 min, followed by 30 cycles of denaturation at 90°C for 30 sec., followed by annealing (REP, 40°C for 1 min; ERIC, 52°C for 1 min), and then extension (65°C for 8 min), with a single final extension (65°C for 16 min). All PCR reaction tubes were placed in internal rows of the thermal cycler and all peripheral tubes were surrounded by "dummy" tubes containing water and mineral oil. Five to eight  $\mu$ l of each PCR reaction volume were then electrophoresed directly on 1% agarose gels containing 1x TAE (Tris acetate-EDTA), and 0.5  $\mu$ g/ml ethidium bromide. The gels were photographed with 20 second exposures to Polaroid type 55 film.

#### Example 6

##### REP Primers

Genomic DNA from lysed *E. coli* W3110 cells served as the test sample. REP1R-I (SEQ. ID. NO. 4) and REP2-I (SEQ. ID. NO. 7) oligonucleotides were used as the pair of outwardly-directed primers. PCR amplification was accomplished as described above. Separation of amplification products was accomplished on 1% agarose - 1x Tris-acetate-EDTA gel, and the pattern of sized extension products was determined using ethidium bromide to stain the DNA. No template DNA was added to the negative control lanes. REP1R-I and REP2-I primers were used in negative control lane 11. Results are shown in Figure 4. The inosine-containing outwardly-directed primer pair, REP1R-I and REP2-I, provided the most distinct genomic fingerprint of *E. coli* strain W3110 chromosomal

-25-

DNA. REP oligonucleotides (Fig. 2) were all tested as primers for DNA amplification because these outwardly-directed primers can amplify DNA between successive REP sequences in any orientation. The inosine-containing primer, however, provided more distinct DNA amplification band patterns and less smearing, possibly because each primer is represented by a single primer sequence instead of a pool of multiple primer sequences as with REP1R-D (SEQ. ID. NO. 5) and REP2-D (SEQ. ID. NO. 8). Each REP primer alone yielded visible amplification products of relatively limited complexity. This result probably stems from the fact that each side of the inverted repeat has a slightly different consensus sequence. The use of both primers REP1R-I and REP2-I appears to allow optimal annealing with both sides of the conserved stem of each REP-like sequence in the genome. Inefficient amplification with REPALL-I (SEQ. ID. NO. 2) and REPALL-D (SEQ. ID. NO. 3) was observed presumably because a palindrome is present in the primer. Because of the possible self-hybridization between REPALL primers of opposite orientation it was not possible to design the primers to the complete REP consensus sequence in both orientations.

## Example 7

## ERIC Primers

Using genomic DNA from lysed *E. coli* W3110 cells as the test sample, ERIC1R (SEQ. ID. NO. 38) and ERIC2 (SEQ. ID. NO. 42) oligonucleotides were used as the pair of outwardly-directed primers. PCR amplification was accomplished as described above. Separation of amplification products was accomplished on 1% agarose - 1x Tris-acetate-EDTA gel, and the pattern of sized extension products was determined using ethidium bromide to stain the DNA. No template DNA was added to the negative control lanes. ERIC1R and ERIC2 primers were used in negative control lane 15. Results are shown in Figure 4. Amplification

-26-

5 results obtained with the single consensus ERIC primer set, ERIC1R and ERIC2 (Fig. 4), were matched in complexity by the results obtained with ERIC2 alone (Fig. 4). In contrast PCR amplification with ERIC1R alone yielded limited amplification products (Fig. 4). Two possible reasons for this observation are that either greater sequence conservation exists in the side of the inverted repeat complementary to ERIC2 or homologous, unrelated sequences complementary to ERIC2 exist outside ERIC elements in the genome.

#### Example 8

##### 10 Specificity of Primer/Template Interactions.

PCR reactions using primer binding sites at known distances from ERIC sequences were performed to verify the size of amplification products. Specificity of REP primer/template interactions was demonstrated by amplification between a known REP sequence and a Tn5  
15 insertion in the *glpD* gene of *E. coli*. The ERIC primers generated PCR products of the expected sizes after amplification of Kohara lambda phages containing the *E. coli hsdR* locus and an adjacent ERIC sequence. Results are shown in Fig. 5. The Kohara lambda phages used are listed by clone numbers and miniset serial numbers are shown in parentheses. One  $\mu$ l of  
20 each Kohara phage lysate was used as template DNA. PCR conditions were as described above. Lanes 2-5 represent PCR amplifications with primers within the *hsdR* gene, hsdR+2758 (SEQ. ID. NO. 58) and hsdR-3235R (SEQ. ID. NO. 59). Lanes 7-10 represent PCR products generated by primers hsdR+2758 and ERIC2. Lane 6 is a blank lane where nothing  
25 was added to the gel. The molecular weight marker is a 1-kb ladder. The gels were 1% agarose-1x Tris-acetate-EDTA and contained 0.5  $\mu$ g of ethidium bromide per ml. The specificity of ERIC-PCR was confirmed by PCR amplification of a defined DNA segment between a published ERIC sequence, (a/k/a as IRU sequence) and a sequence within the *E. coli hsdR*

-27-

gene using the ordered Kohara phage library. Single PCR products of the expected size were amplified both within the *hsdR* gene and between the *hsdR* and ERIC sequences carried by Kohara phages containing the *E. coli* *hsdR* locus. Amplification with only a single *hsdR* primer failed to yield any product.

### Example 9

#### Distinguishing Between Strains of Bacteria With REP

REP primers were used to distinguish different strains within Gram-negative enterobacterial species. Fig. 6 shows extension products generated by PCR amplification of enterobacterial genomic DNA with the REP primers, REP1R-I (SEQ. ID. NO. 4) and REP2-I (SEQ. ID. NO.7). PCR reactions were performed as described above. No template DNA was added to the negative control lane. The DNA molecular weight marker is a 1-kb ladder. The gels were 1% agarose-1x Tris-acetate-EDTA and contained 0.5 µg of ethidium bromide per ml.

The REP-PCR genomic fingerprint of different strains from several bacterial species revealed distinct patterns as shown in Fig. 6. PCR amplification of DNA from multiple strains of different enterobacterial species using primers REP1R-I and REP2-I demonstrated subspecies or strain-specific band patterns. Strains within a species could be unambiguously identified. In lanes 2 and 3 (Fig. 6), *E. coli* K-12 strains HB101 and W3110 were distinguished clearly by an extra band of approximately 400 bp in W3110. Laboratory strains of *E. coli* K-12 were related to each other and distinct from the pathogenic strains of *E. coli*. Interestingly the *Salmonella typhimurium* laboratory stain LT-2 revealed a close similarity to *Salmonella typhi* strain 2304. Both of these strains showed REP-PCR patterns clearly distinct from other pathogenic *Salmonella* strains of undetermined species. The two *Klebsiella pneumoniae* strains shown were obtained from different sources and

-28-

showed different banding patterns. In lanes 14-15 and lanes 20-21 identical strains of pathogenic *Salmonella* and *Enterobacter sakazakii* respectively were represented by identical REP-PCR patterns.

#### Example 10

##### 5 Distinguishing Between Strains of Bacteria With ERIC

ERIC primers were used to distinguish different strains within Gram-negative enterobacterial species. Fig. 7 shows extension products generated by PCR amplification of enterobacterial genomic DNA with the ERIC oligonucleotide primers, ERIC1R (SEQ. ID. NO. 38) and ERIC2 (SEQ. ID. NO. 42) (Fig. 2). PCR reactions were performed as described above. No template DNA was added to the negative control lane. The DNA molecular weight marker is a 1-kb ladder. The gels were 1% agarose-1x Tris-acetate-EDTA and contained 0.5 µg of ethidium bromide per ml.

15 The ERIC-PCR genomic fingerprint of different strains from several bacterial species revealed distinct patterns as shown in Fig. 7. PCR amplification of DNA from multiple strains of different enterobacterial species using primers ERIC1R and ERIC2 demonstrated species specific band patterns. The complexity, however, was less than that obtained with REP-PCR (Fig. 6) and the differences between species were easier to distinguish. This decrease in complexity of the genomic fingerprints, however, made it more difficult to make fine distinctions between strains, for example *E. coli* laboratory strains HB101 and W3110. Greater ERIC-PCR pattern differences existed when comparing laboratory strains of *E.* coli to pathogenic strains of the same species than between laboratory *E.* coli to pathogenic *Shigella*. The ERIC-PCR patterns of greatest complexity were observed with *Salmonella* and these results are consistent with previous data base searches revealing an abundance of ERIC in the *Salmonella*. Because both REP and ERIC PCR yielded common bands

between the strains of a given species it provides the ability to group strains within a certain species.

### Example 11

#### Evolutionary Conservation of REP Sequences

5 Figs. 8 and 9 show the use of REP primers to demonstrate the evolutionary conservation of REP sequences. In Fig. 8 is a listing of bacterial and non-bacterial species which match the genomic DNA in each slot of the bug blot hybridization presented in Fig. 8. The bug blot represents a slot blot DNA:DNA hybridization of genomic DNA probed with <sup>32</sup>P-end-labeled REPALL-D (SEQ. ID. NO. 3). Filters were prepared and hybridizations were performed as described above. Fig. 9 shows two  
10 gels of PCR amplification products of bacterial genomic DNAs used in the bug blot hybridization with REP primers, REP1R-I and REP2-I. These PCR reactions are presented in exactly the same order as the slots of the bug blot. All PCR reactions were performed as described above. No  
15 template DNA was added to the negative control lane. The DNA molecular weight marker is a 1-kb ladder. The gels were 1% agarose-1x Tris-acetate-EDTA and contained 0.5 µg of ethidium bromide per ml.

Slot blot hybridization of the bug blot with SEQ. ID. NO. 3 (Fig. 8)  
20 indicates that Gram-negative enterics and related species from the same phyla comprise the majority of REP-positive species. Hybridizations with REPALL-I and REP2-I yielded results similar to the hybridization with REPALL-D. The 38 mer REPALL probes were used for hybridization because the increased length provides a longer homologous stretch and  
25 hence greater stability for hybridization. As expected several species of Gram-positive bacteria and spirochetes in addition to the phylogenetically distant eukaryotic fungi failed to yield hybridization signals. Surprisingly, however, hybridization signals were observed with distantly related radioresistant bacteria *Deinococcus radiophilus*, the green non-sulfur

bacterium, *Herpetosiphon giganteus*, and the archaebacterium, *Halobacterium halobium*.

PCR amplification of these same bacterial species with primers REP1R-I and REP2-I yielded results consistent with the bug blot hybridization described above. The species that showed the most intense hybridization signals in Fig. 8 generally demonstrated the most complex amplification patterns by REP-PCR (Fig. 9). PCR amplification of genomic DNA from different species clearly revealed species-specific REP-PCR patterns (Fig. 9).

#### Example 12

##### Evolutionary Conservation of ERIC Sequences

Figs. 10 and 11 show the use of ERIC primers to demonstrate the evolutionary conservation of ERIC sequences. In Fig. 10 is a listing of bacterial and non-bacterial species which match the genomic DNA in each slot of the bug blot hybridization presented in Fig. 10. The bug blot represents a slot blot DNA:DNA hybridization of genomic DNA probed with <sup>32</sup>P-end-labeled ERIC2. Filters were prepared and hybridizations were performed as described above. Fig. 11 shows two gels of PCR amplification products from bacterial genomic DNAs used in the bug blot hybridization with the ERIC primers, ERIC1R and ERIC2. These PCR reactions are presented in exactly the same order as the slots of the bug blot. All PCR reactions were performed as described above. No template DNA was added to the negative control lane. The DNA molecular weight marker is a 1-kb ladder. The gels were 1% agarose-1x Tris-acetate-EDTA and contained 0.5 µg/ml of ethidium bromide.

The ERIC primers showed similarity of hybridization and PCR amplification. It should be noted that hybridization with ERICALL yielded results consistent with hybridization with ERIC2. Gram negative enterics and related species from the same family comprised a majority of



-31-

ERIC positive species and as suspected several species of gram positive bacteria and spirochetes in addition to the fungi failed to yield hybridization signals. Similarly to REP (Ex. 11) the radioresistant, green non-sulfur bacterium and the archaebacterium yielded hybridization signals.

ERIC-PCR also provided results (Fig. 11) consistent with ERIC hybridization of the bug blot (Fig. 10). Gram-negative enteric species yielded the amplification patterns of greatest complexity (Fig. 11). Most Gram-positive species (e.g. *Bacillus subtilis*) showed minimal ERIC-PCR amplification (Fig. 11). This result is consistent with computer searches of ERIC in DNA sequence databases and known phylogenetic distances between Gram-positive bacteria and Gram-negative enteric bacteria.

#### Example 13

##### Bacterial DNA Fingerprint Library

The method described above was used to screen a plurality of different bacterial strains. The pattern for each strain was categorized and stored. This comprehensive library of fingerprints was used to compare with unknown samples to determine the strain identity.

#### Example 14

##### Whole Cell PCR

Gram-negative bacterial colonies are picked with disposable loops, and the cells on the edge of the loops are placed directly into PCR tubes containing PCR reaction buffer. The repetitive sequence oligonucleotide primers are then added with dNTP's and DNA polymerase, and the PCR reactions are carried out. Presumably during the initial denaturation step at 94°C the cells lyse and the chromosomal DNA released into solution

-32-

serves as the template for PCR amplification. Thus DNA isolation and purification prior to PCR amplification is not always necessary.

### Example 15

#### Genome Mapping

5            REP-PCR was performed on purified cosmid DNA from the ordered  
Tabata cosmid library. Tabata, et al., J Bacteriol., 171:1214-1218 (1989).  
This library covers approximately 70% of the *E. coli* strain W3110  
genome. This library represents a set of overlapping or isolated cosmids  
10        which contain genomic DNA from different locations on the *E. coli*  
chromosome. Each individual cosmid DNA is purified and used as  
template DNA in the PCR reaction. 100 ng DNA of each cosmid DNA  
(represented by serial numbers in Figs. 14-15) is used as template in REP-  
PCR with primers REP1R-I and REP2-I (50 pmol of each primer). The  
15        PCR products are then electrophoresed in 1% agarose, 1x TAE, and  
stained with 0.5 micrograms per ml. ethidium bromide.

As is evident from Figs. 14-15, the different cosmids have different  
REP-PCR fingerprints, depending on which segment of the genome is  
inserted into a particular clone. By matching fingerprint patterns from  
individual clones with the computer, contiguous (contigs) and ordered  
20        stretches of overlapping clones are built.

Further, this fingerprinting method provides a useful tool for  
checking the integrity of the library and the purity of each clone. One  
skilled in the art will readily recognize that these libraries can be made  
mfrom cosmid, phage, or any possible DNA (even RNA) vector.

-33-

## Example 16

## Fingerprinting Bacterial

## Strains Used in Newborn Screening

The techniques of the present invention were used to monitor the validity of bacterial strains used in newborn screening. REP-PCR, ERIC-PCR, and combined REP/ERIC - PCR were performed on *Bacillus subtilis* strains, ATCC 6633 and 6051, which are used for newborn screening of phenylketonuria (PKU) and maple-syrup urine disease (MSUD) respectively. In the Fig. 13, REP1R-I, REP2-I, ERIC1R, and ERIC2 (50 pmol of each primer) were used in single PCR reactions (REP-ERIC - PCR) on individual samples of *Bacillus subtilis* genomic DNA. One strain that was supposedly ATCC 6633 turned out to be an anomalous strain (lane 12) that was clearly distinct from the others. The strain used for MSUD diagnosis, ATCC 6051, was distinguished from the strain used for PKU diagnosis, ATCC 6633. No template DNA was added to the negative control lane. PCR products were electrophoresed on 1% agarose gels in 1x TAE and stained with 0.5 micrograms per ml. ethidium bromide. It should also be noted that this example shows the combination of two sets of different primers and their simultaneous use in identifying and fingerprinting strains of bacteria.

## Example 17

## Ngrep

In Fig. 12 are the results of PCR using Ngrep primers (SEQ. ID. NOS. 48 and 52). As can be readily seen, the different strains of *Neisseria* can be distinguished. The conditions are as described in previous examples except the denaturing and annealing steps occur at 94°C for 1 min. and 38°C for 1 min., respectively. The negative control has no DNA template but includes primers.

-34-

5 All patents and publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

10 One skilled in the art will readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well those inherent therein. The outwardly-directed primers, along with the methods and procedures described herein are presently representative of preferred embodiments, are exemplary, and not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention or defined by the scope of the claims.

-35-

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Lupski, James R.  
Versalovic, James  
Koeuth, Thearith

5

(ii) TITLE OF INVENTION: Fingerprinting Bacterial Strains Using  
Repetitive DNA Sequence Amplification

(iii) NUMBER OF SEQUENCES: 60

10

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Fulbright & Jaworski  
(B) STREET: 1301 McKinney, Suite 5100  
(C) CITY: Houston  
(D) STATE: Texas  
(E) COUNTRY: U.S.A.  
(F) ZIP: 77010-3095

15

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: Patent In Release #1.0, Version #1.25

20

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US  
(B) FILING DATE:  
(C) CLASSIFICATION:

25

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Paul, Thomas D.  
(B) REGISTRATION NUMBER: 32,714  
(C) REFERENCE/DOCKET NUMBER: D-5394

30

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 713/651-5325  
(B) TELEFAX: 713/651-5246

-36-

(C) TELEX: 762829

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCCGATGNC GRCGYNNNNN RCGYCTTATC MGGCCTAC

38

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

-37-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCCNGATGNC GNCGNNNNNN NCGNCTTATC NGGCCTAC

38

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 38 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCCXGATGNC GRCGYNNNNN RCGYCTTATC MGGCCTAC

38

15

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= N = "Inosine"

-38-

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

NNNNCGNCGN CATCGGC

18

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine at position #1-3; N = A, G, C or T at position #10"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

15 NNNRCGYCGN CATCMGGC

18

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES



-39-

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine at position #1-3; N = A, G, C or T at position #4, 7, 10, 15"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

5

NNNNCGNCGN CATCNGGC

18

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

15

(A) OTHER INFORMATION: /note= "N = Inosine"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

NCGNCTTATC NGGCCTAC

18

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

WO 93/08297

-40-

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

18

RCGYCTTATC MGGCCTAC

(2) INFORMATION FOR SEQ ID NO:9:

5

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

18

15

NCGNCTTATC NGGCCTAC

(2) INFORMATION FOR SEQ ID NO:10:

20

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

-41-

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

10

GNCATCNGGC

5 (2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

10

GNCATCMGGC

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

-42-

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GNCATCNGGC

10

5

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

NCGNCATCNG GC

12

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

-43-

(iii) HYPOTHETICAL: YES.

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

5

YCGNCATCMG GC

12

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

NCGNCATCNG GC

12

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

-44-

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

NCGNNNNNNN CGNCGNCATC NGGC

24

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

15

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

RCGYNNNNNR CGYCGNCATC MGGC

24

(2) INFORMATION FOR SEQ ID NO:18:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

-45-

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

5

(A) OTHER INFORMATION: /note= "N = Inosine at position #5-9; N = A, G, C or T at position #1, 4, 10, 13, 16, 21"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

NCGNNNNNNN CGNCGNCATC NGGC

24

(2) INFORMATION FOR SEQ ID NO:19:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

20

ATAAGNCGNN NNNNNCGNCG NCATCNGGC

29

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

-46-

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

10 ATAAGRCGYN NNNNRCGYCG NCATCMGGC

29

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

20 (A) OTHER INFORMATION: /note= "N = Inosine at position  
#10-14; N = A, G, C or T at position #6, 9, 15, 18, 21, 16"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATAAGNCGNN NNNNRCGNCG NCATCNGGC

29



-47-

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCNGGCCTAC

10

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCMGGCCTAC

10

## (2) INFORMATION FOR SEQ ID NO:24:

WO 93/08297

-48-

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 10 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

10

TCNGGCCTAC

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 12 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

20

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TATCNGGCCT AC

12

-49-

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: YES

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

10 TATCMGGCCT AC

12

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: YES

## (ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = A, G, C or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TATCNGGCCT AC

12

## (2) INFORMATION FOR SEQ ID NO:28:

-50-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = Inosine"

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

NNNNNNNCGN CTTATCNGGC CTAC

24

## (2) INFORMATION FOR SEQ ID NO:29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

20

## (ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

YNNNNNNRCGY CTTATCMGGC CTAC

24

-51-

## (2) INFORMATION FOR SEQ ID NO:30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine at position #2-6; N = A, G, C or T at position #1, 7, 10, 17"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

NNNNNNNCGN CTTATCNGGC CTAC

24

## (2) INFORMATION FOR SEQ ID NO:31:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

-52-

CGNCGNNNNN NNCGNCTTAT CNGGCCTAC

29

## (2) INFORMATION FOR SEQ ID NO:32:

## (i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

10

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGRCGYNNNN NRCGYCTTAT CMGGCCTAC

29

## (2) INFORMATION FOR SEQ ID NO:33:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine at position  
#7-11; N = A, G, C or T at position #3, 6, 12, 15, 22"

-53-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CGNCGNNNNN NNCGNCTTAT CNGGCCTAC

29

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

NNNNACGCCG CATCCGGC

18

15

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

-54-

18

TCGGCTTATC GGGCCTAC

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 126 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = A, G, C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TATACMCWAA ATHATTTCGRG TTGCAKSAAG GCGGCAASNK AGTGAATYCC CRGGAGCWTA 60

SATAASTAWG TGACTGGGRT GARCRRRCGM AGCCAACGCA SMTGCRYYYY GAARKAYGAM 120

GRGKAT

126

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 126 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES



-55-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TATACCCAAA ATAATTCGAG TTGCAGCAAG GCGGCAAGTG AGTGAATCCC CAGGAGCTTA 60  
CATAAGTAAG TGACTGGGGT GAGCGAACGC AGCCAACGCA GCTGCAGCTT GAAATATGAC 120  
GGGTAT 126

5 (2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATGTAAGCTC CTGGGGATTC AC 22

15 (2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = Inosine"

-56-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

22

ATNTANGCTC CNGGGNATTC AC

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

22

ATSTAWGCTC CYGGGRATTC AC

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

-57-

ATNTANGCTC CNGGGNATTC AC

22

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AAGTAAGTGA CTGGGGTGAG CG

22

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AANTANGTGA CTGGGNTGAN C

21

-58-

## (2) INFORMATION FOR SEQ ID NO:44:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

10 AASTAWGTGA CTGGGRTGAR C

21

## (2) INFORMATION FOR SEQ ID NO:45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = A, G, C or T"

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AANTANGTGA CTGGGNTGAN C

21

## (2) INFORMATION FOR SEQ ID NO:46:

-59-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

## (ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T at all locations; and at location 8 the N can be omitted to form a 26 mer sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GTNCNGNNTT TTTGTTAATN CNCTATA

27

(2) INFORMATION FOR SEQ ID NO:47:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTACCGGTTT TTTGTTAATTC ACTATA

26

(2) INFORMATION FOR SEQ ID NO:48:

-60-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

ACAAAAACCG GTAC

14

10

(2) INFORMATION FOR SEQ ID NO:49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = Inosine"

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

ACAAAAANCN GNAC

14

(2) INFORMATION FOR SEQ ID NO:50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs

-61-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

ACAAAAAYCR GKAC

14

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iii) HYPOTHETICAL: YES

(ix) FEATURE:

- (A) OTHER INFORMATION: /note= "N = A, G, C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

ACAAAAANCN GNAC

14

20 (2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-62-

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

5 GTTAATTCAC TATA

14

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = Inosine"

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GTTAATNCNC TATA

14

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20



-63-

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GTTAATYCRC TATA

14

5 (2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(ix) FEATURE:

(A) OTHER INFORMATION: /note= "N = A, G, C or T"

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GTTAATNCNC TATA

14

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

20 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

-64-

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CGAGCTGTCC CAGTCCGC

18

(2) INFORMATION FOR SEQ ID NO:57:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GCGGACTGGG ACAGCTCG

18

(2) INFORMATION FOR SEQ ID NO:58:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CAGCCATGAA CAACTGGTGG CG

22

-65-

## (2) INFORMATION FOR SEQ ID NO:59:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

10 TGCTTTGCGC AGGGAAGATT CC

22

## (2) INFORMATION FOR SEQ ID NO:60

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 90 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

20 YTAGAGYATT TGMCAAAAAG ACGCAACGTC TTTTGGCCG GCGGACTGGG ACAGCTCGMA 60  
GAGRCCGAGT GCAAAACACK GAGCAGGGCG 90

-66-

CLAIMS

What we claim is:

1. A method for identifying a strain of bacteria, comprising the steps of:
  - 5 amplifying DNA between repetitive sequences in a sample containing said bacteria by adding a pair of outwardly-directed primers to said sample, said primers capable of hybridizing to repetitive DNA sequences in the bacterial DNA and extending outwardly from one hybridizable repetitive sequence to another hybridizable repetitive sequence;
  - 10 separating the extension products generated in the amplification step by size; and
  - 15 determining the specific strain of bacteria by measuring the pattern of sized extension products.
2. The method of claim 1 wherein the hybridizable repetitive sequence is selected from the group consisting of repetitive extragenic palindromic elements (REP), enterobacterial repetitive intergenic consensus sequence (ERIC), Neisseria repetitive extragenic elements (Ngrep), Deinococcus repetitive extragenic elements (Drrep) and any combination thereof.
- 20 3. The method of claim 1, wherein the primers are between about 10 mer and 29 mer.
4. The method of claim 1 wherein the primers are 15 mer to 25 mer.
- 25 5. The method of claim 1 wherein the repetitive sequence is REP and one member of the pair of primers is selected from the group consisting of SEQ. ID. Nos. 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 34; and the other member of the pair is selected from the group consisting of SEQ. ID. NOS. 7, 8, 9, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33 and 35.
- 30

-67-

6. The method of claim 1, wherein the repetitive sequence is ERIC and one member of the pair of primers is selected from the group consisting of SEQ. ID. NOS. 38, 39, 40 and 41; and the other member of the pair is selected from the group consisting of SEQ.  
5 ID. NOS. 42, 43, 44 and 45.
7. The method of claim 1, wherein the repetitive sequence is Ngrep and one member of the pair of primers is selected from the group consisting of SEQ. ID. NOS. 48, 49, 50 and 51; and the other member of the pair is selected from the group consisting of 52, 53,  
10 54 and 55.
8. The method of claim 1, wherein the primers are SEQ. ID. NO. 4 and SEQ. ID. NO. 7.
9. The method of claim 1, wherein the repetitive sequence is ERIC and the primers are SEQ. ID. NO. 38 and SEQ. ID. NO. 42.
- 15 10. The method of claim 1, wherein the repetitive sequence is Ngrep and the primers are SEQ. ID. NO. 48 and SEQ. ID. NO. 52.
11. The method of claim 1, wherein the repetitive sequence is Drrep and the primer is SEQ. ID. NO. 56 or SEQ. ID. NO. 57.
12. The method of claim 1, wherein a plurality of pairs are added and  
20 wherein each pair is to a different repetitive sequence.
13. The method of claim 12, wherein primers are selected from the group consisting of REP, ERIC, Ngrep, Drrep and any combination thereof.
14. The method of claim 12, wherein the primers are SEQ. ID. NOS.  
25 4, 7, 38 and 42.
15. The method of claim 1, wherein the DNA is extracted from the bacteria prior to adding the primers.
16. The method of claim 1, wherein the separating step includes gel electrophoresis of the extension products.

17. The method of claim 16, wherein the extension products are stained with ethidium bromide.
18. The method of claim 1 wherein the primers are labelled and the determining step includes measuring the pattern of labelling.
- 5 19. The method of claim 1 wherein the separation step includes chromatography of the extension products.
20. The method of claim 19, wherein the primers are labelled.
21. The method of claim 19, wherein the label is a fluorescer.
22. The method of claim 1, wherein the sample contains a plurality of  
10 bacteria and wherein each specific strain of bacteria is distinguished by its unique size pattern of extension products.
23. The method of claim 22, wherein the sample is selected from the group consisting of blood, urine, spinal fluid, tissue, vaginal swab, stool, amniotic fluid, and buccal mouthwash.
- 15 24. The method of claim 23, wherein the sample is from a human subject.
25. The method of claim 23, wherein the test sample is from an animal subject.
26. The method of claim 22, wherein the test sample is an agriculture  
20 sample.
27. The method of claim 22, wherein the test sample is food.
28. The method of claim 22, where in the test sample is an environmental sample.
29. The method of claim 22, wherein the test sample is a horticulture  
25 sample.
30. The method of claim 1 for diagnosing bacterial disease wherein the sample is collected from a subject suspected of having a bacterial disease.
31. The method of claim 30, wherein the subject is a human.
- 30 32. The method of claim 30, wherein the subject is an animal.

-69-

33. The method of claim 30, wherein the subject is a plant.
34. The method of claim 1 for monitoring bacterial contamination in an environment wherein the sample is collected from an environmental source suspected of being contaminated;
- 5 35. The method of claim 34, wherein the environmental source is a liquid.
36. The method of claim 34, wherein the environmental source is sludge.
37. The method of claim 34, wherein the environmental source is sewage.
- 10 38. The method of claim 34, wherein the environmental source is a treatment plant.
39. The method of claim 34, wherein the environmental source is soil.
40. The method of claim 1 for monitoring bacterial contamination of food wherein the sample is collected from food suspected of being contaminated.
- 15 41. The method of claim 40, wherein the food is infant formula.
42. The method of claim 40, wherein the food is sea food.
43. The method of claim 40, wherein the food is fresh produce.
- 20 44. The method of claim 40, wherein the food is processed food.
45. The method of claim 1 for monitoring a bacterial population at a bioremediation site wherein the sample is collected from said site.
46. The method of claim 45, wherein the sample is soil.
47. The method of claim 45, wherein the sample is liquid.
- 25 48. The method of claim 45, wherein the sample is sludge.
49. The method of claim 45, wherein the sample is from the bacteria which is to be added to the site.
50. The method of claim 1 for monitoring a horticulture sample wherein the sample is collected from a horticulture source to be tested.
- 30

51. The method of claim 1 for monitoring an agriculture sample wherein the sample is collected from an agriculture source to be tested.
52. The method of claim 1 for monitoring bacterial additions to an agricultural environment wherein the sample is collected from an agriculture source to be tested.
53. The method of claim 52, wherein the sample is a liquid.
54. The method of claim 52, wherein the sample is soil.
55. The method of claim 52, wherein the sample is from a plant.
56. The method of claim 52, wherein the sample is from an animal.
57. The method of claim 1 for monitoring manufacturing processes for bacteria wherein the sample is collected from the process to be tested.
58. The method of claim 57, wherein the sample is selected from the group consisting of drug manufacturing processes, fermentation processes, microorganism-aided synthesis processes, chemical manufacturing process and food manufacturing processes.
59. The method of claim 1 for quality assurance/quality control of laboratory tests involving microbiological assays wherein the sample is collected from the bacterial stock to be tested.
60. The method of claim 1 for tracing outbreaks of bacterial infections, wherein the sample is collected from an organism to be tested.
61. A method for genome mapping, comprising the steps of:  
fractionating the genome;  
cloning the fractionated genome into a vector;  
testing the cloned vectors by amplifying bacterial DNA in the clones by adding a pair of outwardly-directed primers to the test sample, said primers capable of hybridizing to repetitive DNA sequences in the bacterial DNA and extending outwardly from the



-71-

hybridizable repetitive sequence to another hybridizable repetitive sequence;

separating the extension products of the amplification step by size; and

5 measuring the pattern of extension products; and

reconstructing the genome from the overlapping patterns.

62. A method for automated identification of a strain of bacteria comprising the steps of:

10 adding bacteria and outwardly-directed PCR primers to a test sample in an auto-PCR instrument, wherein said primers are capable of hybridizing to repetitive DNA sequence in the bacterial DNA and extending outwardly from the hybridizable repetitive sequence to another hybridizable repetitive sequence.

15 transferring the extension products from the PCR assay and separating the extension products;

measuring the sizing pattern of said separated extension products with a measuring means; and

recognizing and identifying the sizing pattern by a computer means.

20 63. The method of claim 62, wherein the measuring means is selected from the group consisting of a bar code reader, a laser reader, a digitizer, a photometer and a fluorescence reader.

64. The method of claim 62 wherein the extension products are separated by chromatography or gel electrophoresis.

25 65. The method of 62, wherein a sample from the PCR amplification is applied to a gel and electrophoresed; the size pattern is read by a measuring means; and the pattern is compared by the computer means with stored known bacterial patterns.

30 66. The method of 62, wherein the separated extension products are stained with ethidium bromide before reading.

-72-

67. The method of 62, wherein the primers are labelled.
68. The method of 62, wherein the primers are labelled with fluorescer.
69. A method of identifying a strain of bacteria in a test sample, comprising the steps of:
- 5           amplifying DNA in said bacteria by adding a plurality of pairs of outwardly-directed primers to said test sample, each pair of said primers capable of hybridizing to different repetitive DNA sequences in the bacterial DNA and each pair extending outwardly from its hybridizable repetitive sequence to another of its hybridizable repetitive sequences and wherein each pair is
- 10           differentially labelled;
- separating the extension products generated in the amplification step by size;
- determining the specific strain of bacteria by measuring the
- 15           pattern of sized extension products for each pair.
70. The method of 69, wherein the labels are fluorescers.
71. The method of 69, wherein the separation is by gel electrophoresis, capillary electrophoresis, mass spectrometry or chromatography.
72. The method of claim 69, wherein the primer pairs are selected from
- 20           the group consisting of REP, ERIC, Ngrep, Drrep and any combination thereof.
73. A kit for determining the identity of strains of bacteria, comprising a container having outwardly-directed PCR primer pairs to repetitive sequences in bacteria.
- 25           74. The kit of claim 73, wherein the PCR primer pairs are selected from the group consisting of SEQ. ID. NOS. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 38, 39, 40, 41, 42, 43, 44, 45, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57 and any combination thereof.

-73-

75. As a composition of matter the sequences selected from the group consisting of SEQ. ID. NOS. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 and 33.
- 5 76. As a composition of matter the sequences selected from the group consisting of SEQ. ID. NOS. 38, 39, 40, 41, 42, 43, 44 and 45.
77. As a composition of matter the sequences selected from the group consisting of SEQ. ID. NOS. 48, 49, 50, 51, 52, 53, 54 and 55.
78. As a composition of matter the sequences selected from the group consisting of SEQ. ID. NOS. 56 and 57.
- 10 79. As a composition of matter sequences selected from the group consisting of SEQ. ID. NOS. 4 and 7.
80. As a composition of matter sequences selected from the group consisting of SEQ. ID. NOS. 38 and 42.
- 15 81. As a composition of matter sequences selected from the group consisting of SEQ. ID. NOS. 48 and 52.
82. A machine for identifying a strain of bacteria comprising:  
an automated PCR amplifying means;  
a separation means;  
20 a sampling means for removing the extension products from the PCR means and transferring them to the separation means;  
a reading means for measuring patterns of extension products after separation on the separation means;  
a computer means for recording the results of the reading  
25 means and for outputting the pattern and identifying the strain of bacteria.
83. The machine of claim 82, wherein the separation means is selected from the group consisting of a gel electrophoresis apparatus, capillary electrophoresis apparatus, mass spectrometer and a  
30 chromatographic apparatus.

-74-

84. The method of claim 82 wherein, the reading means is selected from the group consisting of a digitizer, a bar code reader, a laser detector, a fluorescence detector, a photometer and a radioactive detector.
- 5 85. A machine for identifying a strain of bacteria comprising:  
a thermal cyclor;  
a separator selected from the group consisting of a gel electrophoresis, capillary electrophoresis and chromatographic apparatus;  
10 a detector selected from the group consisting of a digitizer, a fluorescence detector and a photometer;  
a robotic apparatus for handling samples and moving them from one location to another; and  
a computer means for regulating the operation of the  
15 machine, for collecting and storing the information and for comparing the genomic fingerprints to identify the bacterial strain; wherein the thermocycler, detector and separator are attached to the computer and integrated into a cabinet.
- 20 86. The method of claim 1, wherein the blood sample is tested for bacterial contamination and the sample is stored blood or blood used for transfusions.
87. The method of claim 86, wherein the sample is tested for the specific bacterial species *Yersinia enterocolitica*.
- 25 88. A method for identifying a strain of bacteria, comprising the steps of:  
amplifying DNA between repetitive sequences in a sample containing said bacteria by adding a primer to said sample, wherein said primer is hybridizable to the repetitive sequence in either of the complementary strands of the DNA and wherein the hybridized  
30 primer extends from one repetitive sequence across non-repetitive

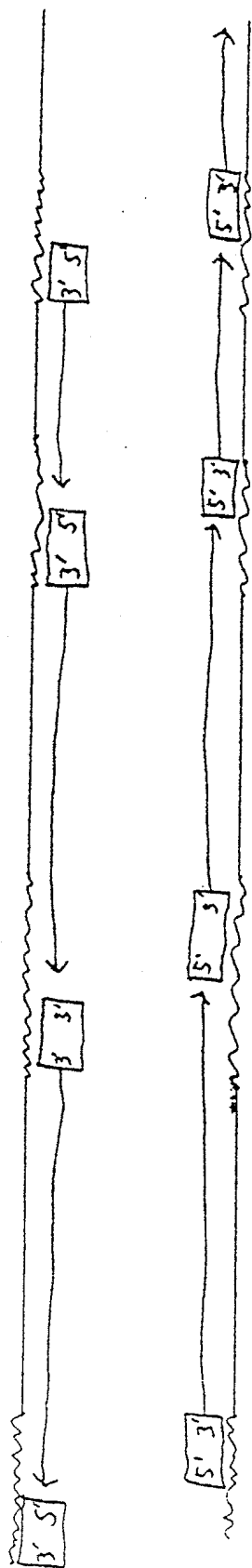
-75-

DNA to another repetitive sequence and wherein said extension product is hybridizable by the primer for generations of further extension products;

separating the extension products by size; and

5 determining the specific strain of bacteria by measuring the pattern of sized extension products.

1/19



— = non-repetitive sequence

~~~~~ = repetitive sequence

[5' 3'] = primer

FIGURE 1

2/19

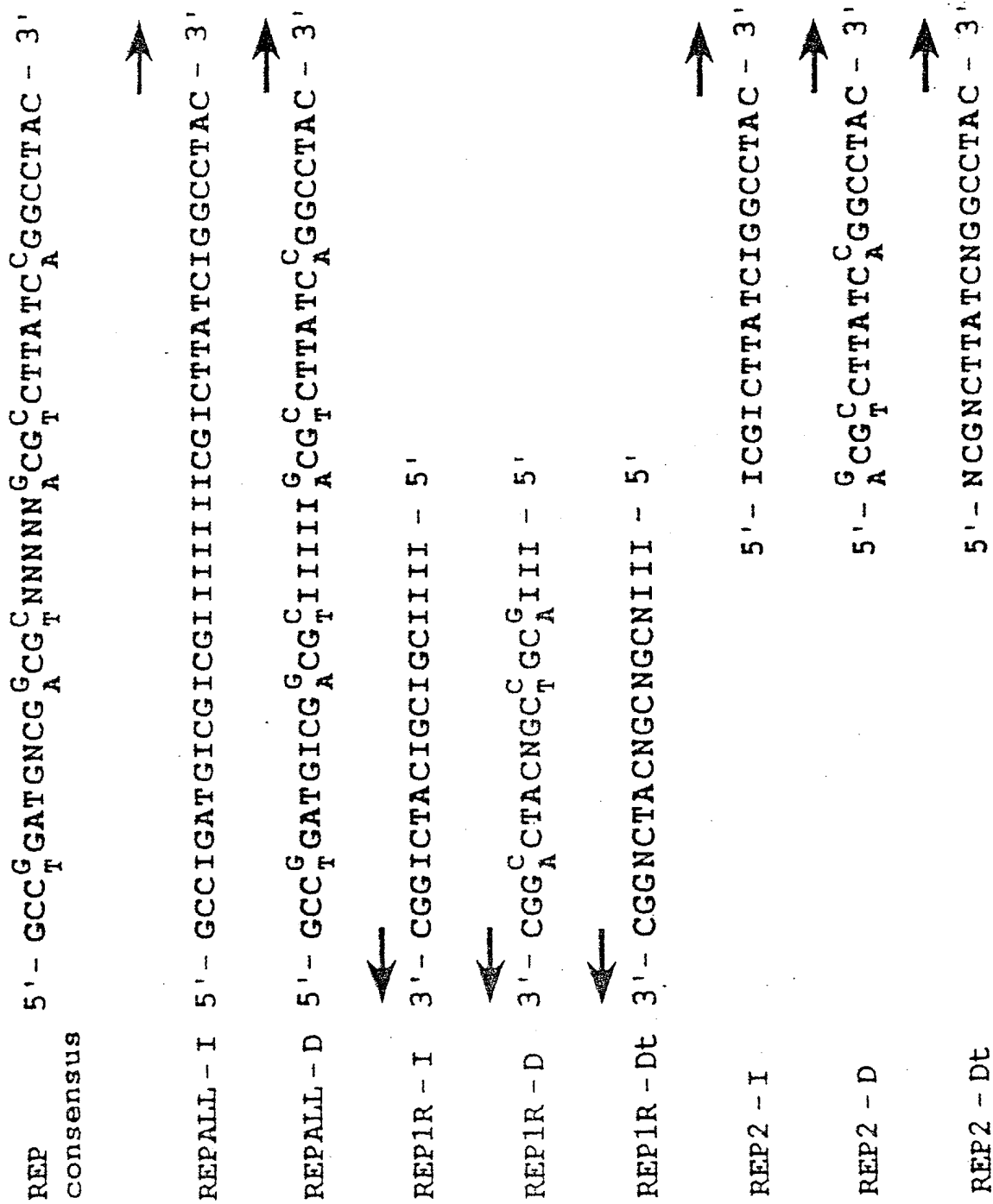


Figure 2

3/19

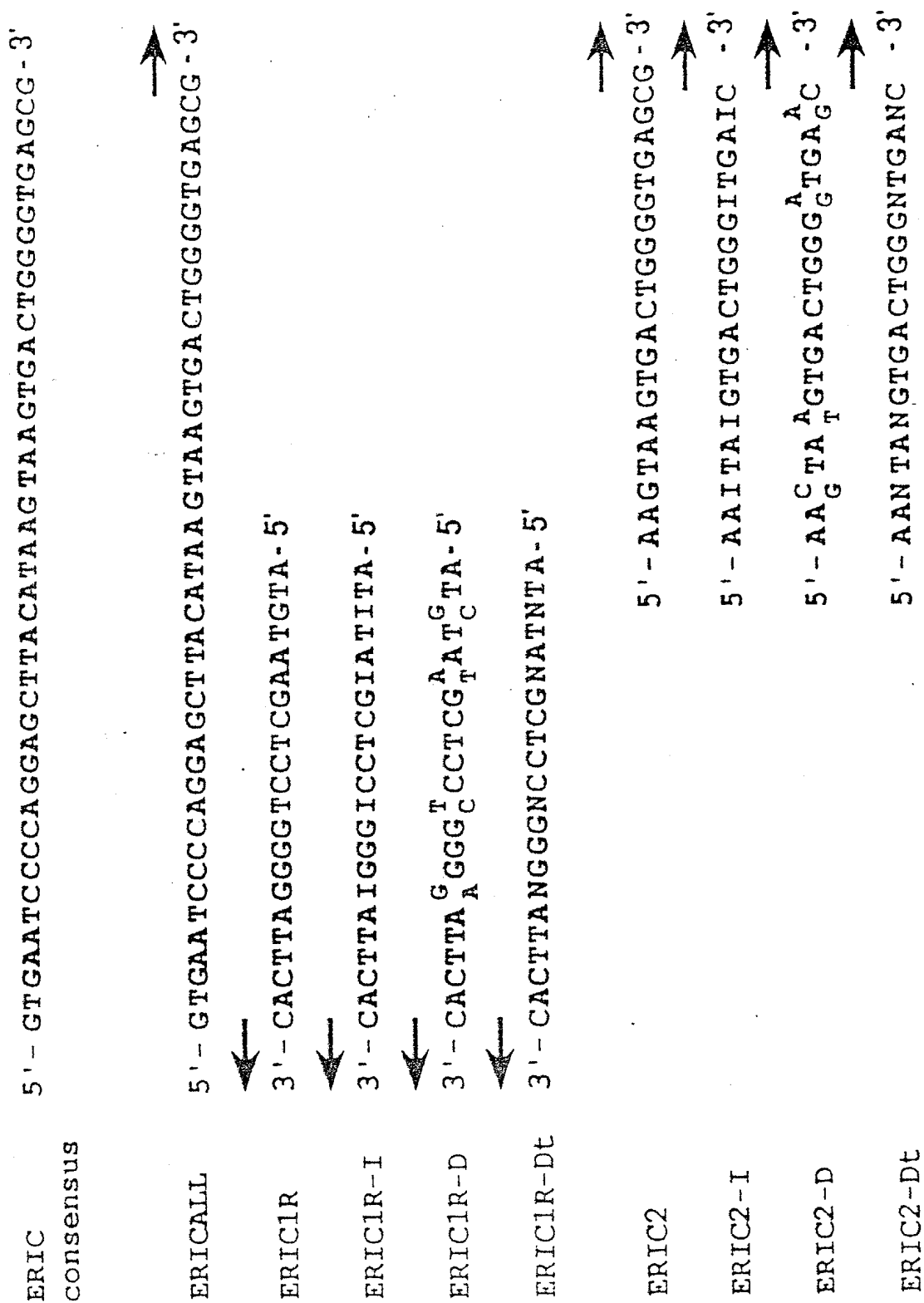


Figure 3



4/19

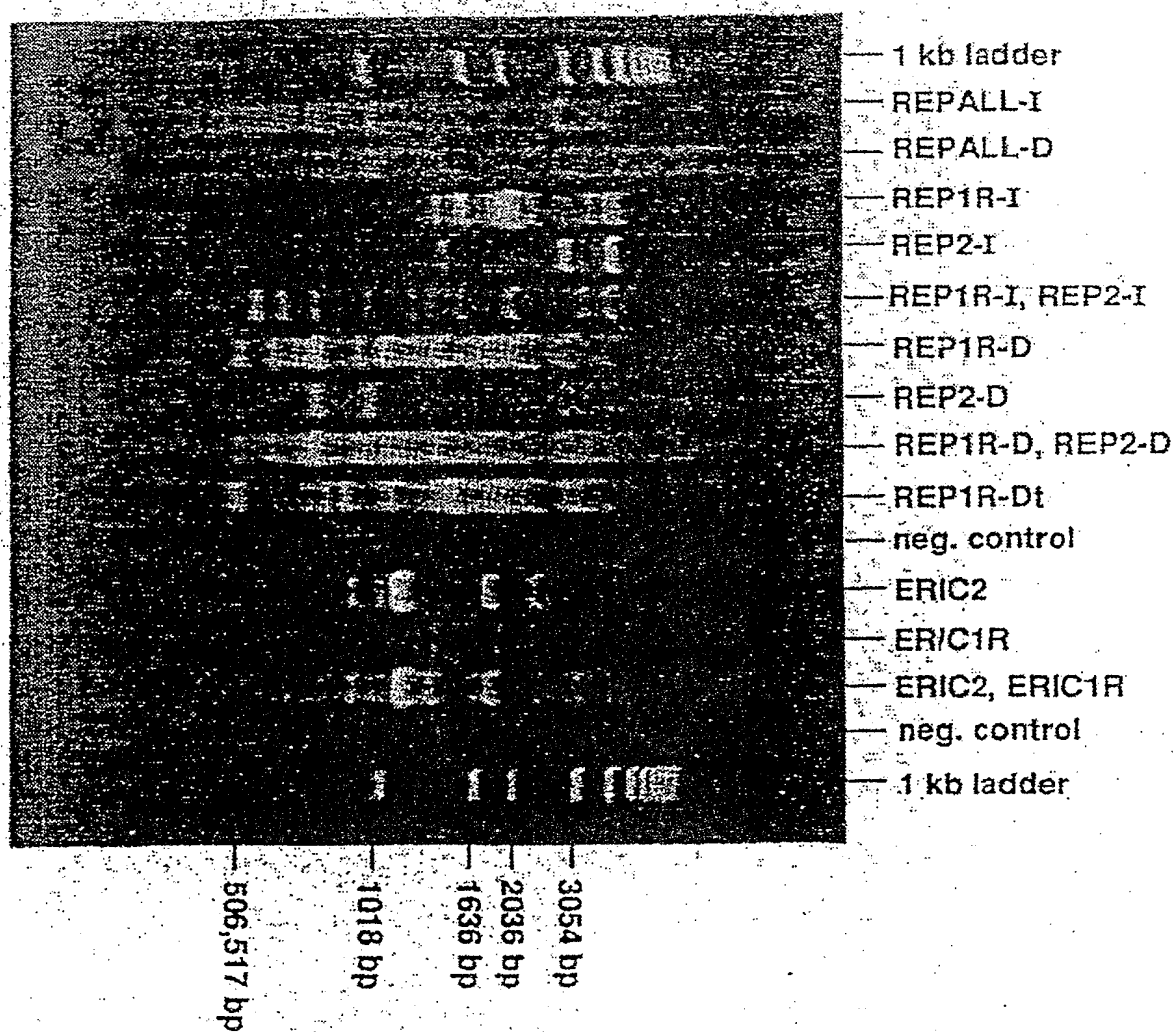


Figure 4

5/19

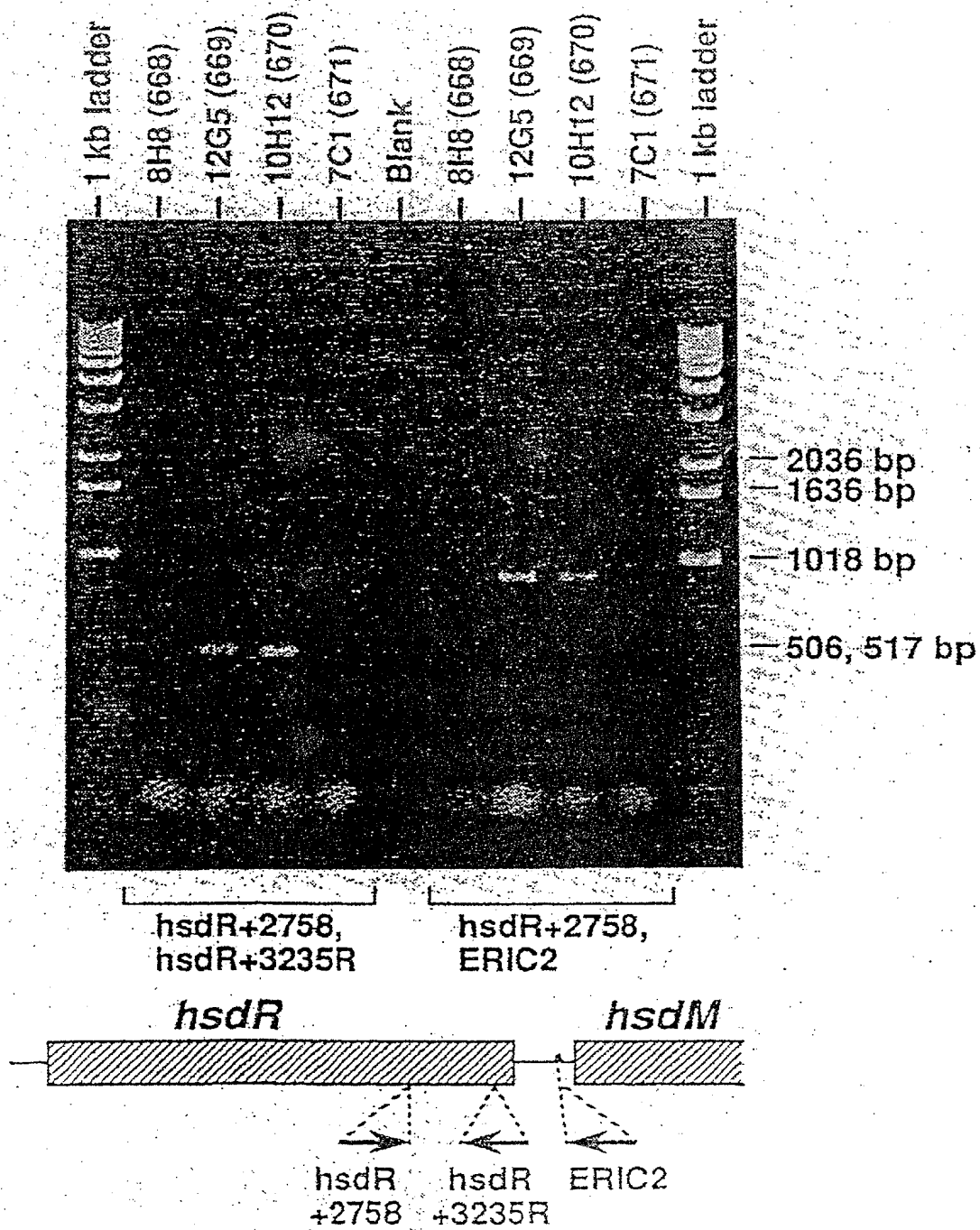


Figure 5

6/19

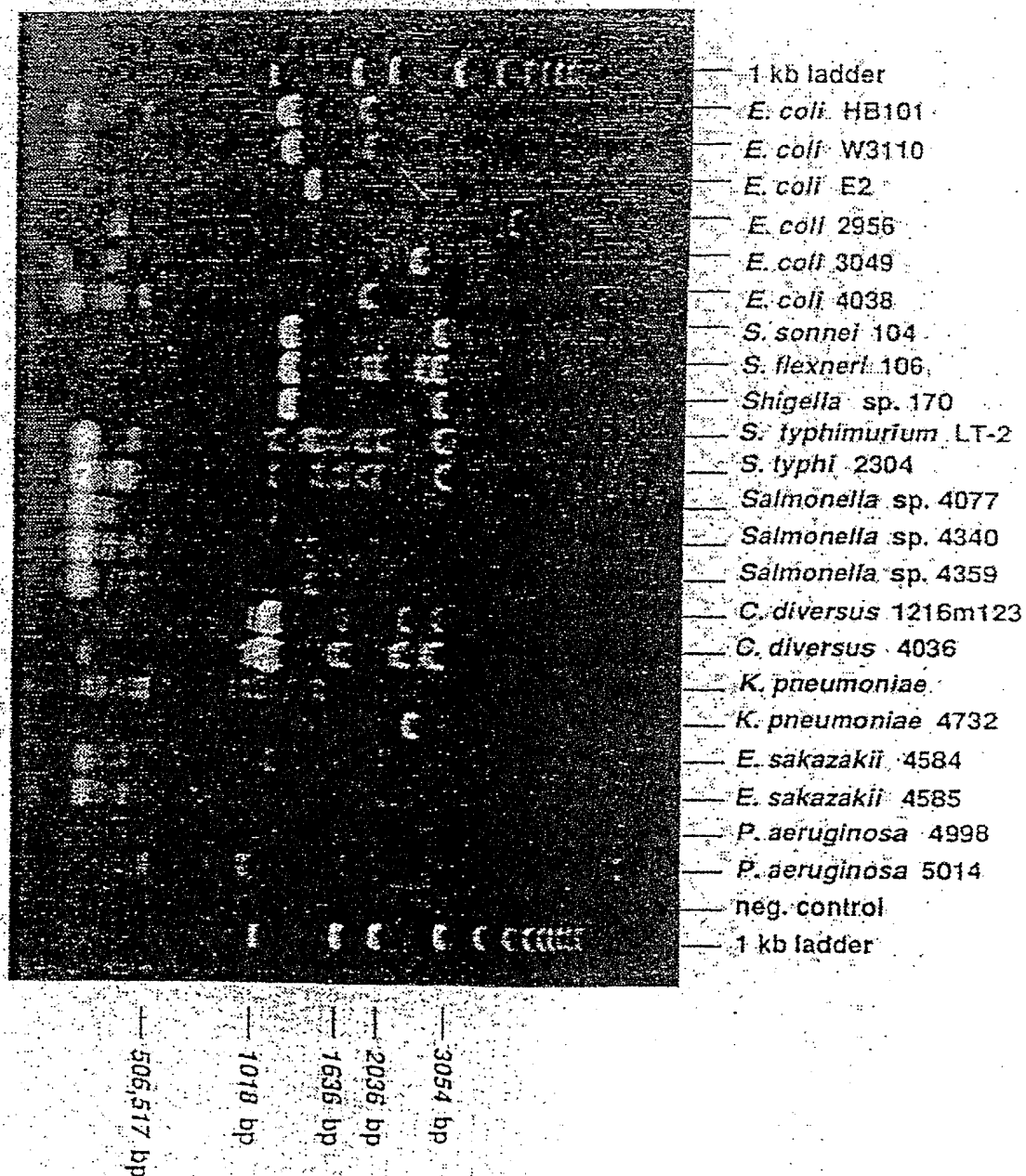


Figure 6

7/19

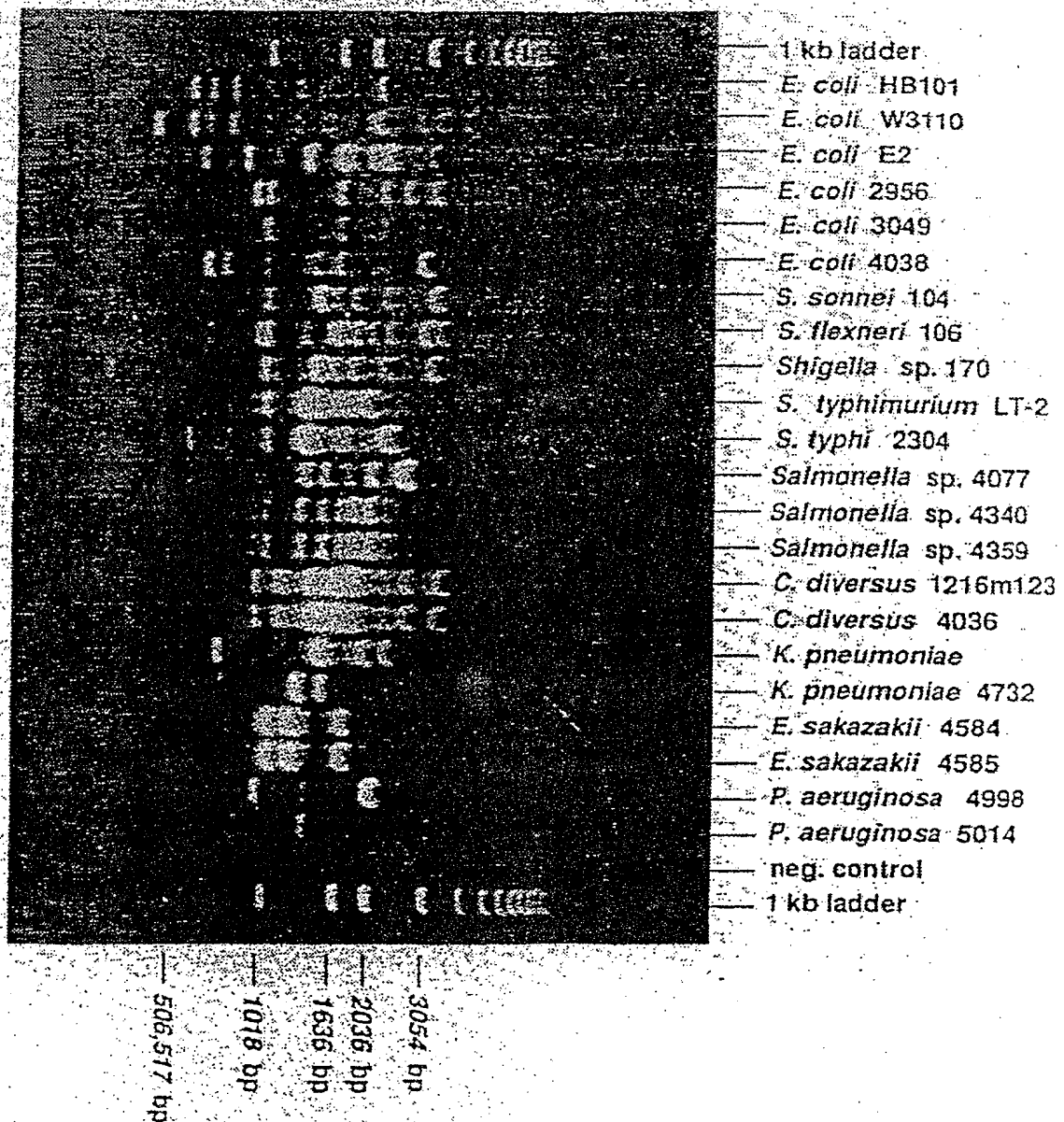


Figure 7

8/19

- |                                                |                                                     |                                                  |
|------------------------------------------------|-----------------------------------------------------|--------------------------------------------------|
| 1. <i>Rhodobacter</i><br><i>sphaeroides</i>    | 2. <i>Rhizobium</i><br><i>meliloti</i>              | 3. <i>Neisseria</i><br><i>gonorrhoeae</i>        |
| 4. <i>Neisseria</i><br><i>meningitidis</i>     | 5. <i>Sphaerotilus</i><br>sp.                       | 6. <i>E. coli</i><br>HB101                       |
| 7. <i>E. coli</i><br>W3110                     | 8. <i>Salmonella</i><br>sp.                         | 9. <i>Citrobacter</i><br><i>diversus</i>         |
| 10. <i>Klebsiella</i><br><i>pneumoniae</i>     | 11. <i>Enterobacter</i><br><i>sakazakii</i>         | 12. <i>Serratia</i><br><i>marcescens</i>         |
| 13. <i>Proteus</i><br><i>vulgaris</i>          | 14. <i>Pseudomonas</i><br><i>aeruginosa</i>         | 15. <i>Xanthomonas</i><br><i>manihotis</i>       |
| 16. <i>Vibrio</i><br><i>vulnificus</i>         | 17. <i>Myxococcus</i><br><i>xanthus</i>             | 18. <i>Arthrobacter</i><br><i>luteus</i>         |
| 19. <i>Nocardia</i><br><i>ottidiscavium</i>    | 20. <i>Streptomyces</i><br><i>albus</i> G           | 21. <i>Mycobacterium</i><br><i>aurum</i>         |
| 22. <i>Bacillus</i><br><i>subtilis</i>         | 23. <i>Listeria</i><br><i>monocytogenes</i>         | 24. <i>Staphylococcus</i><br><i> aureus</i>      |
| 25. <i>Streptococcus</i><br><i>pneumoniae</i>  | 26. Group B<br><i>Streptococcus</i>                 | 27. <i>Caryophanon</i><br><i>latum</i>           |
| 28. <i>Mycoplasma</i><br><i>pneumoniae</i>     | 29. <i>Anabaena</i><br>sp.                          | 30. <i>Borrelia</i><br><i>burgdorferi</i>        |
| 31. <i>Treponema</i><br><i>pallidum</i>        | 32. <i>Treponema</i><br><i>phagedenis</i>           | 33. <i>Bacteroides</i><br><i>fragilis</i>        |
| 34. <i>Fusobacterium</i><br><i>nucleatum</i>   | 35. <i>Flavobacterium</i><br><i>meningosepticum</i> | 36. <i>Flavobacterium</i><br><i>oceanokoites</i> |
| 37. <i>Deinococcus</i><br><i>radiophilus</i>   | 38. <i>Thermus</i><br><i>aquaticus</i>              | 39. <i>Thermus</i><br><i>thermophilus</i>        |
| 40. <i>Herpetosiphon</i><br><i>giganteus</i>   | 41. <i>Halobacterium</i><br><i>halobium</i>         | 42. <i>Saccharomyces</i><br><i>cerevisiae</i>    |
| 43. <i>Schizosaccharomyces</i><br><i>pombe</i> | 44. <i>Candida</i><br><i>parapsilosis</i>           | 45. <i>Homo</i><br><i>sapiens</i>                |

Figure 8A

9/19

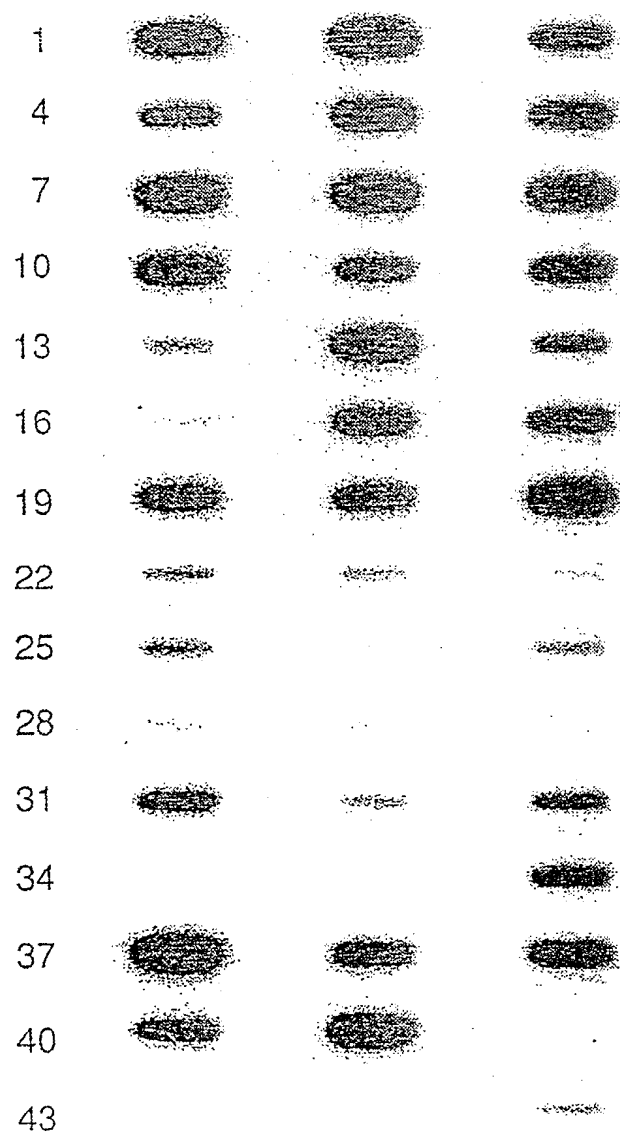


Figure 8B

10/19

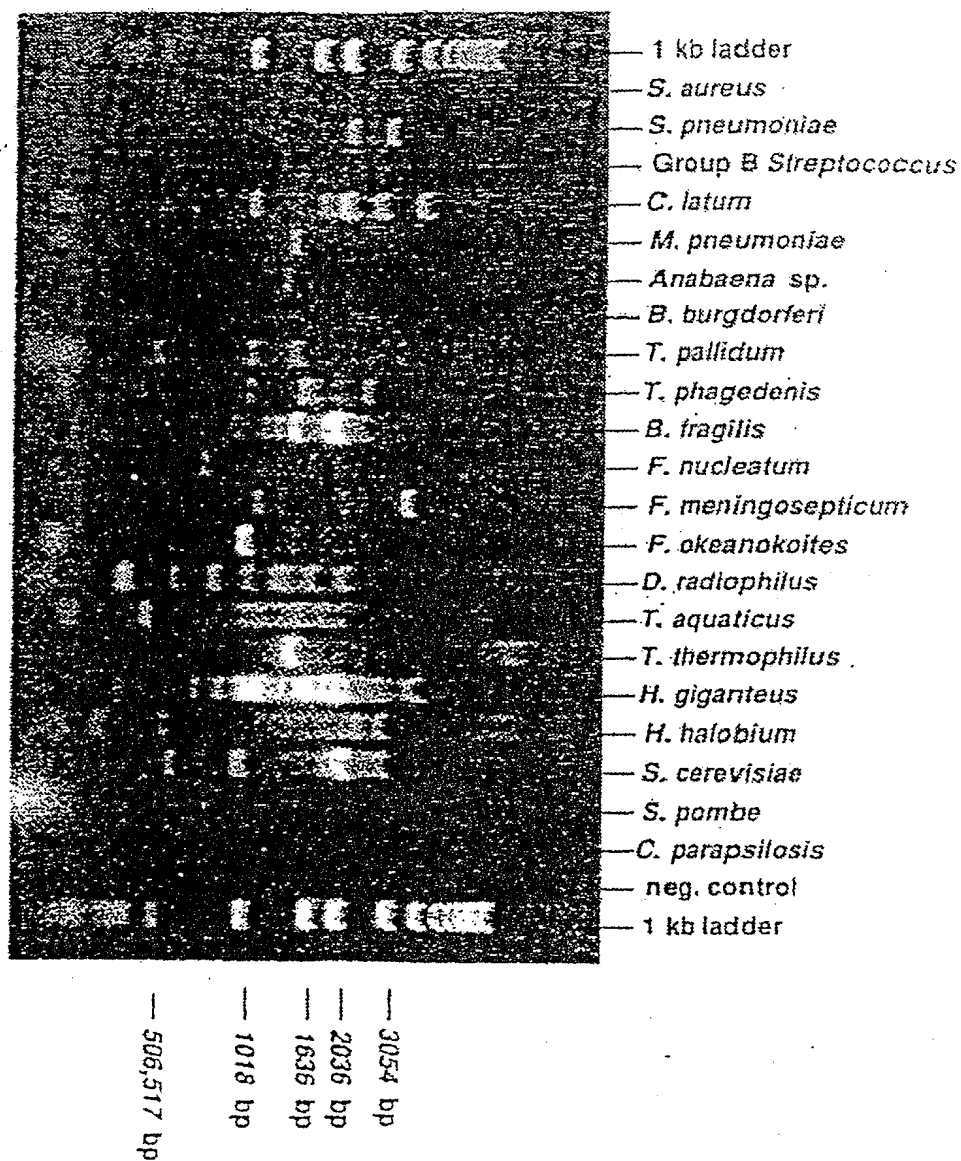


Figure 9A

11/19

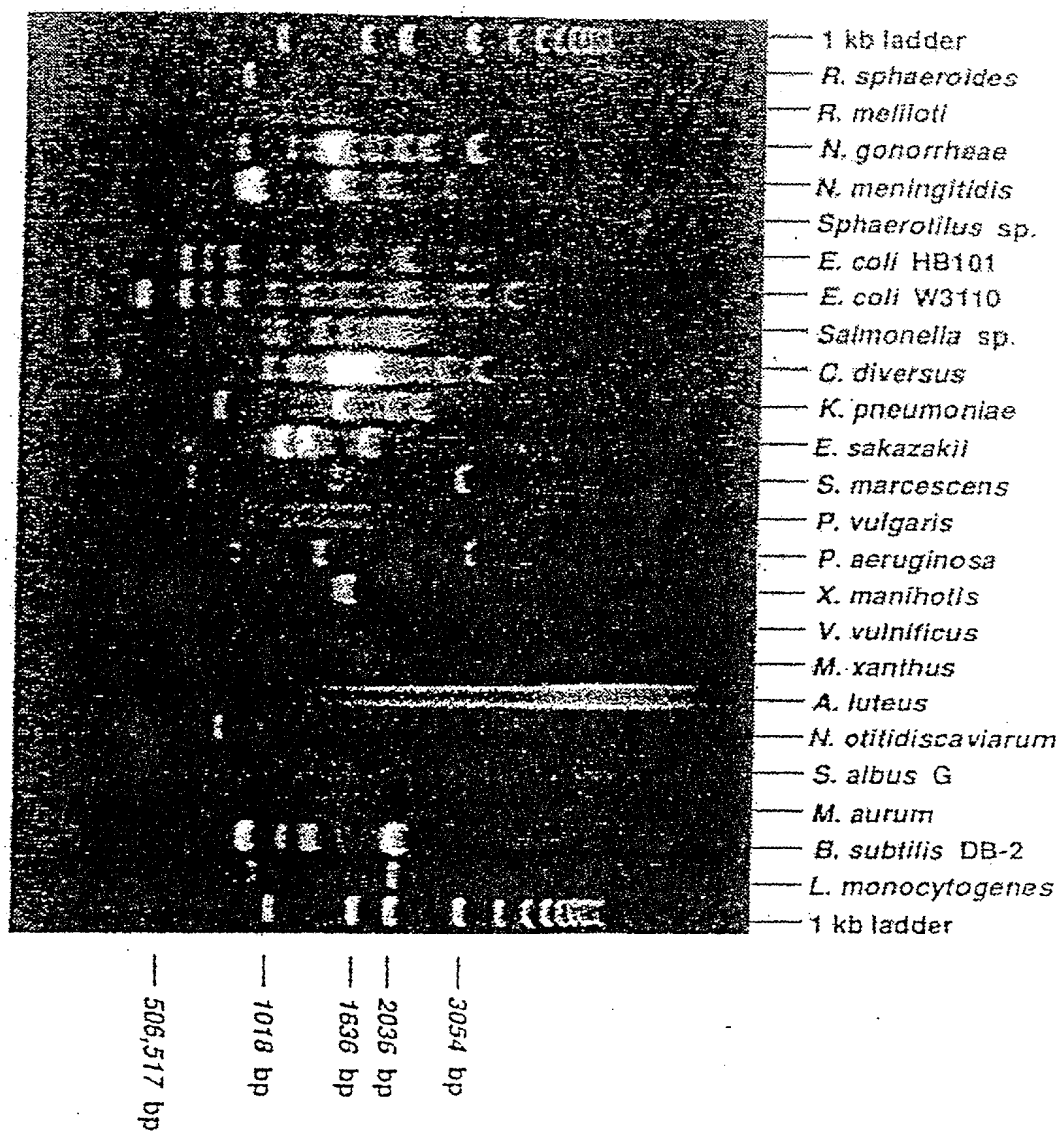


Figure 9B



12/19

- |                                                |                                                     |                                                  |
|------------------------------------------------|-----------------------------------------------------|--------------------------------------------------|
| 1. <i>Rhodobacter</i><br><i>sphaeroides</i>    | 2. <i>Rhizobium</i><br><i>mellotii</i>              | 3. <i>Neisseria</i><br><i>gonorrhoeae</i>        |
| 4. <i>Neisseria</i><br><i>meningitidis</i>     | 5. <i>Sphaerotilus</i><br>sp.                       | 6. <i>E. coli</i><br>HB101                       |
| 7. <i>E. coli</i><br>W3110                     | 8. <i>Salmonella</i><br>sp.                         | 9. <i>Citrobacter</i><br><i>diversus</i>         |
| 10. <i>Klebsiella</i><br><i>pneumoniae</i>     | 11. <i>Enterobacter</i><br><i>sakazakii</i>         | 12. <i>Serratia</i><br><i>marcescens</i>         |
| 13. <i>Proteus</i><br><i>vulgaris</i>          | 14. <i>Pseudomonas</i><br><i>aeruginosa</i>         | 15. <i>Xanthomonas</i><br><i>manihotis</i>       |
| 16. <i>Vibro</i><br><i>vulnificus</i>          | 17. <i>Myxococcus</i><br><i>xanthus</i>             | 18. <i>Arthrobacter</i><br><i>luteus</i>         |
| 19. <i>Nocardia</i><br><i>otitidiscalearum</i> | 20. <i>Streptomyces</i><br><i>albus</i> G           | 21. <i>Mycobacterium</i><br><i>aurum</i>         |
| 22. <i>Bacillus</i><br><i>subtilis</i>         | 23. <i>Listeria</i><br><i>monocytogenes</i>         | 24. <i>Staphylococcus</i><br><i>aureus</i>       |
| 25. <i>Streptococcus</i><br><i>pneumoniae</i>  | 26. Group B<br><i>Streptococcus</i>                 | 27. <i>Caryophanon</i><br><i>latum</i>           |
| 28. <i>Mycoplasma</i><br><i>pneumoniae</i>     | 29. <i>Anabaena</i><br>sp.                          | 30. <i>Borrelia</i><br><i>burgdorferi</i>        |
| 31. <i>Treponema</i><br><i>pallidum</i>        | 32. <i>Treponema</i><br><i>phagedenis</i>           | 33. <i>Bacteroides</i><br><i>fragilis</i>        |
| 34. <i>Fusobacterium</i><br><i>nucleatum</i>   | 35. <i>Flavobacterium</i><br><i>meningosepticum</i> | 36. <i>Flavobacterium</i><br><i>okeanokoites</i> |
| 37. <i>Deinococcus</i><br><i>radiophilus</i>   | 38. <i>Thermus</i><br><i>aquaticus</i>              | 39. <i>Thermus</i><br><i>thermophilus</i>        |
| 40. <i>Herpetosiphon</i><br><i>giganteus</i>   | 41. <i>Halobacterium</i><br><i>halobium</i>         | 42. <i>Saccharomyces</i><br><i>cerevisiae</i>    |
| 43. <i>Schizosaccharomyces</i><br><i>pombe</i> | 44. <i>Candida</i><br><i>parapsilosis</i>           | 45. <i>Homo</i><br><i>sapiens</i>                |

Figure 10A

13/19

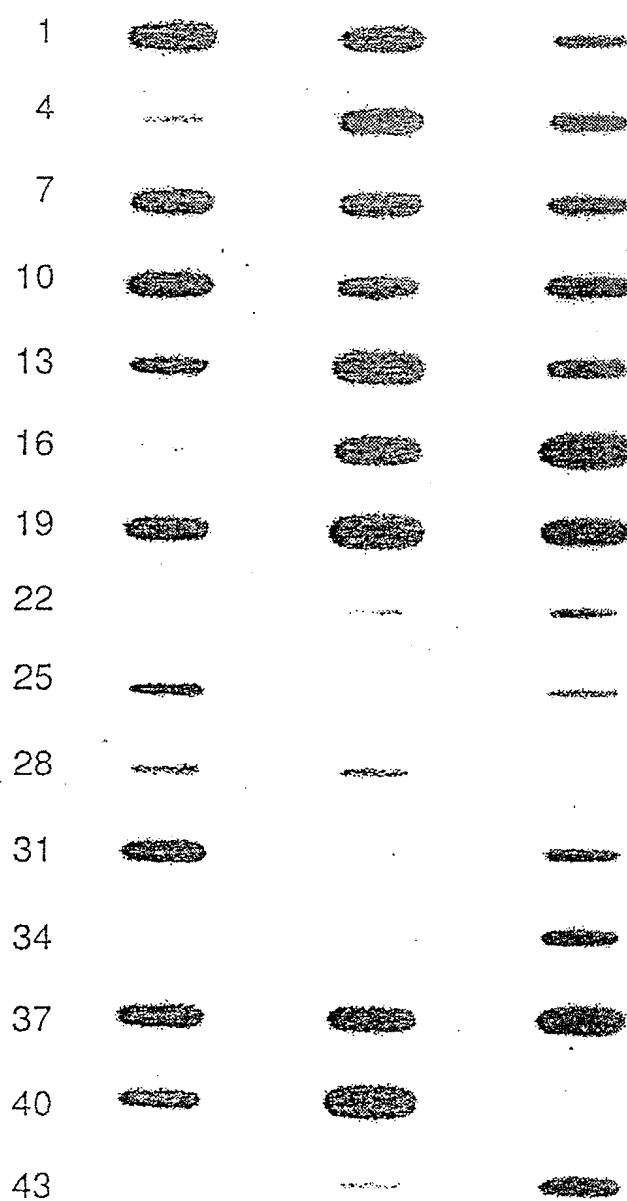


Figure 10B

14/19

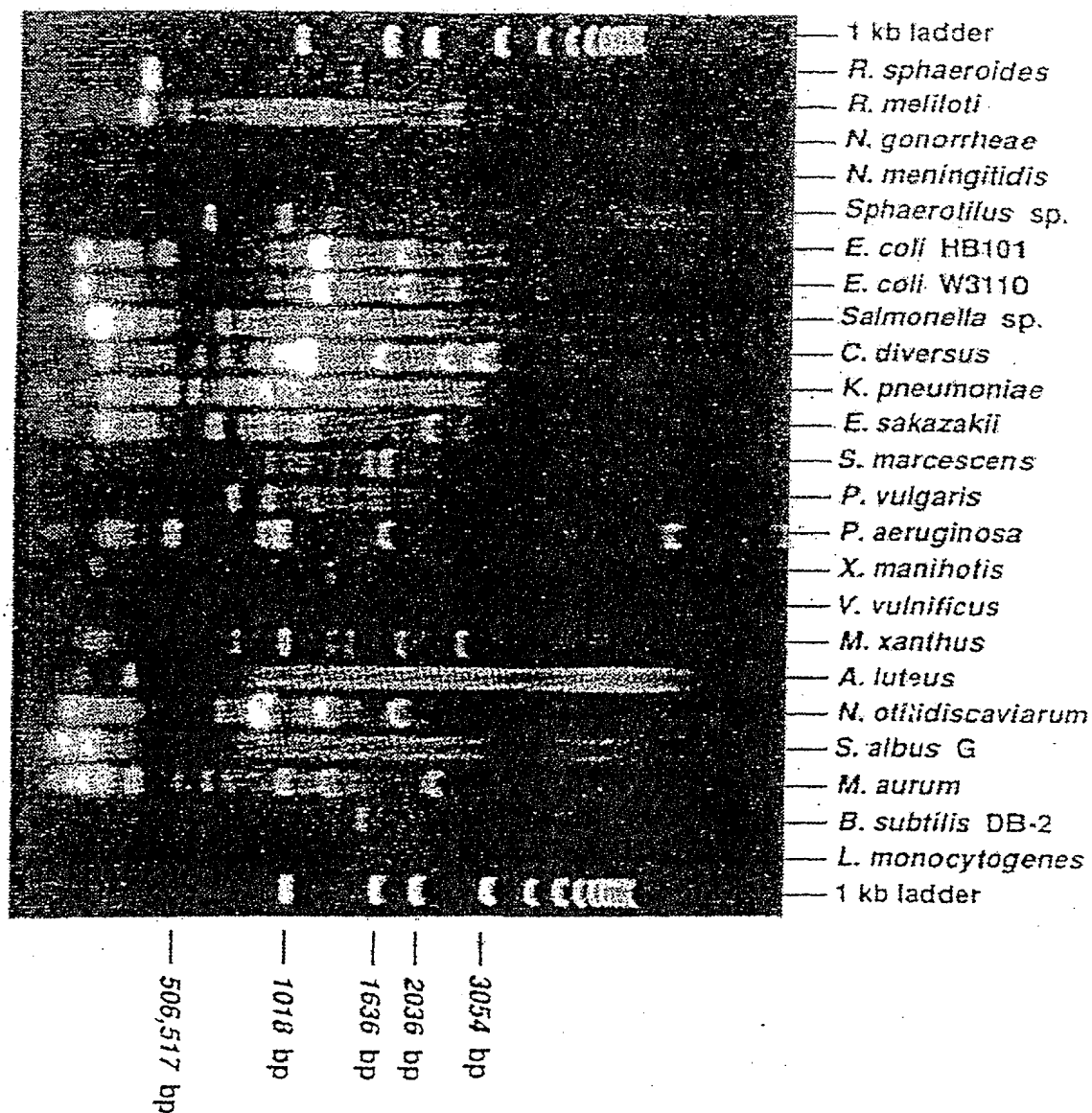


Figure 11A

15/19

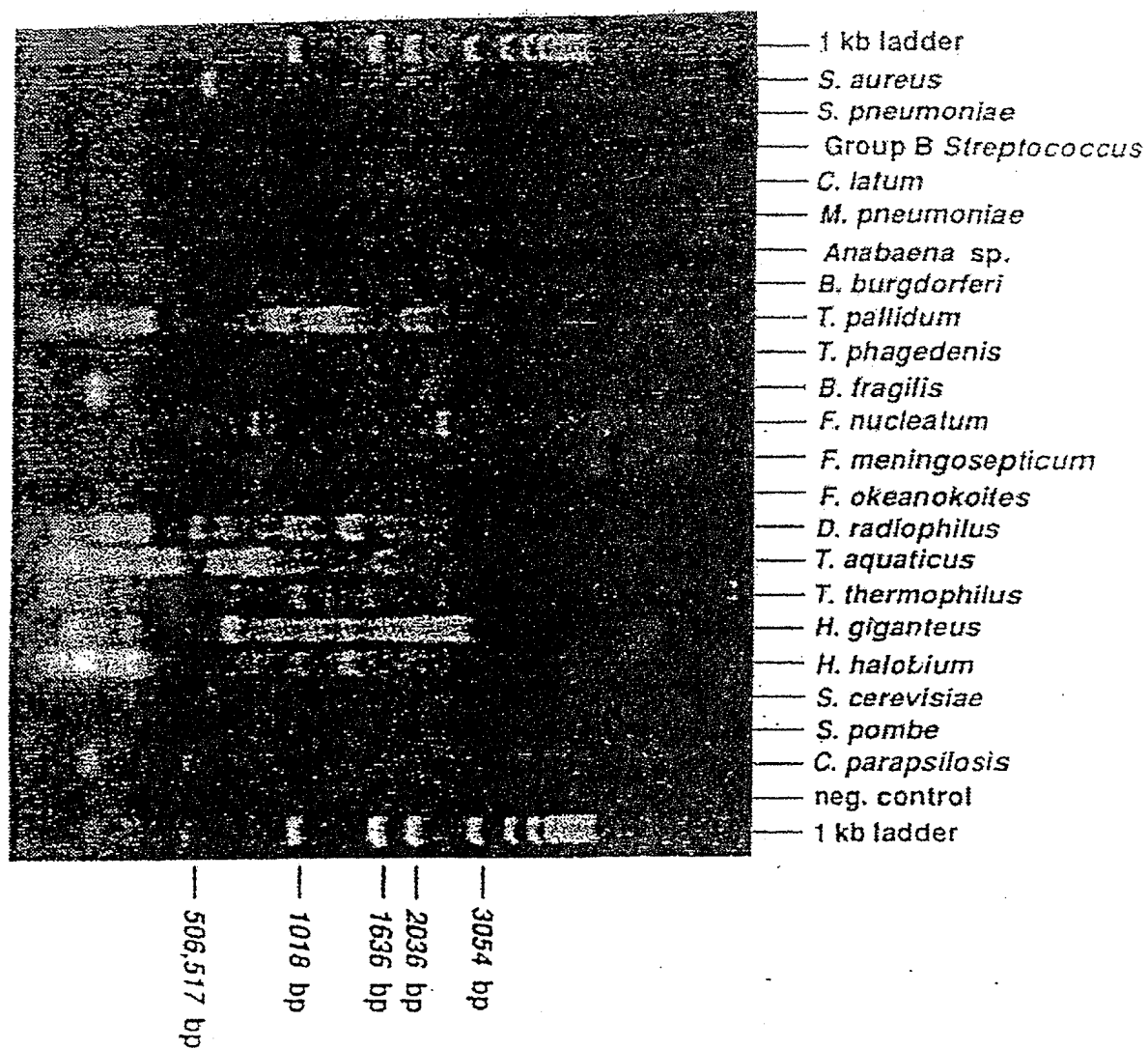


Figure 11B

16/19

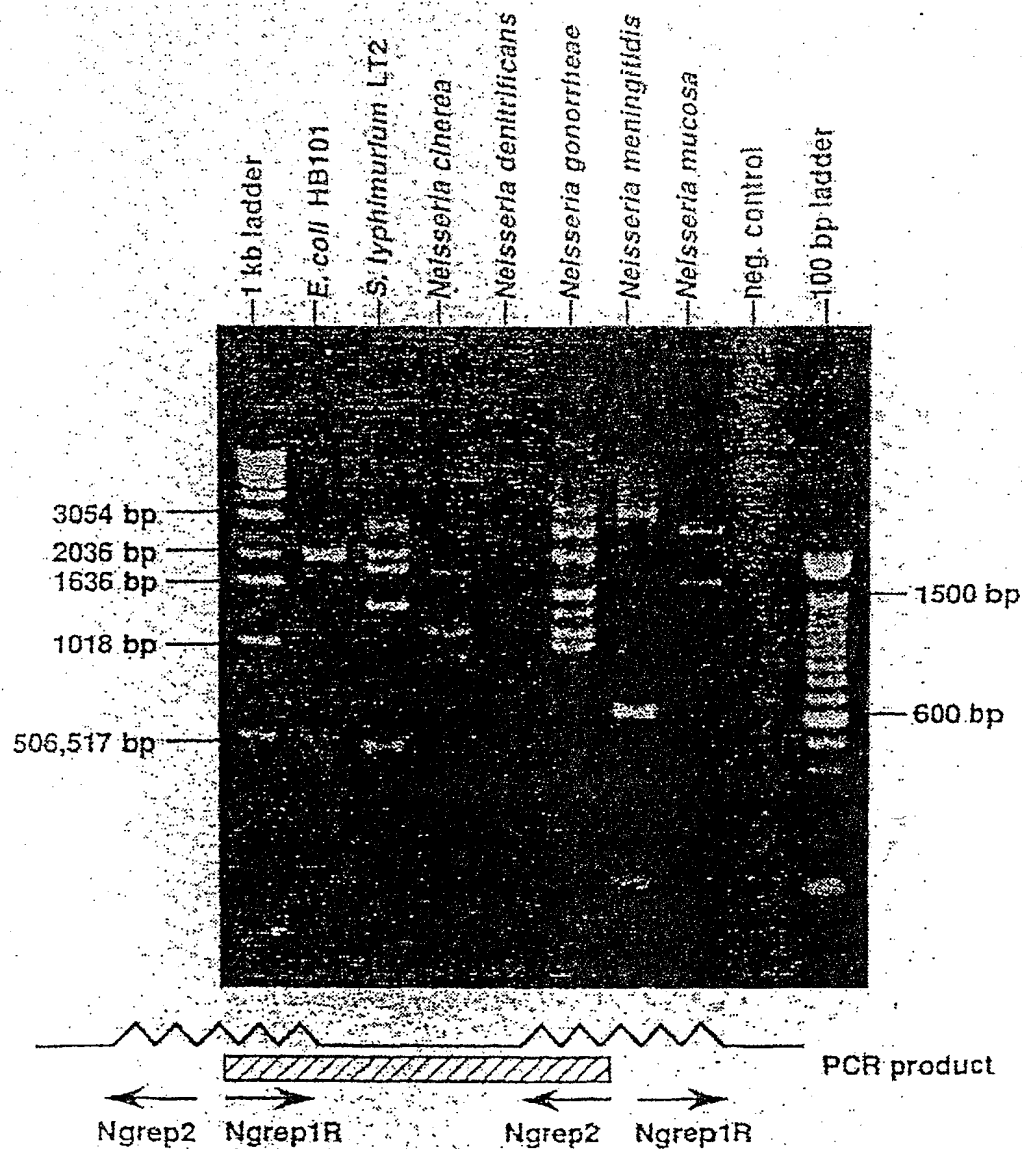


Figure 12

17/19

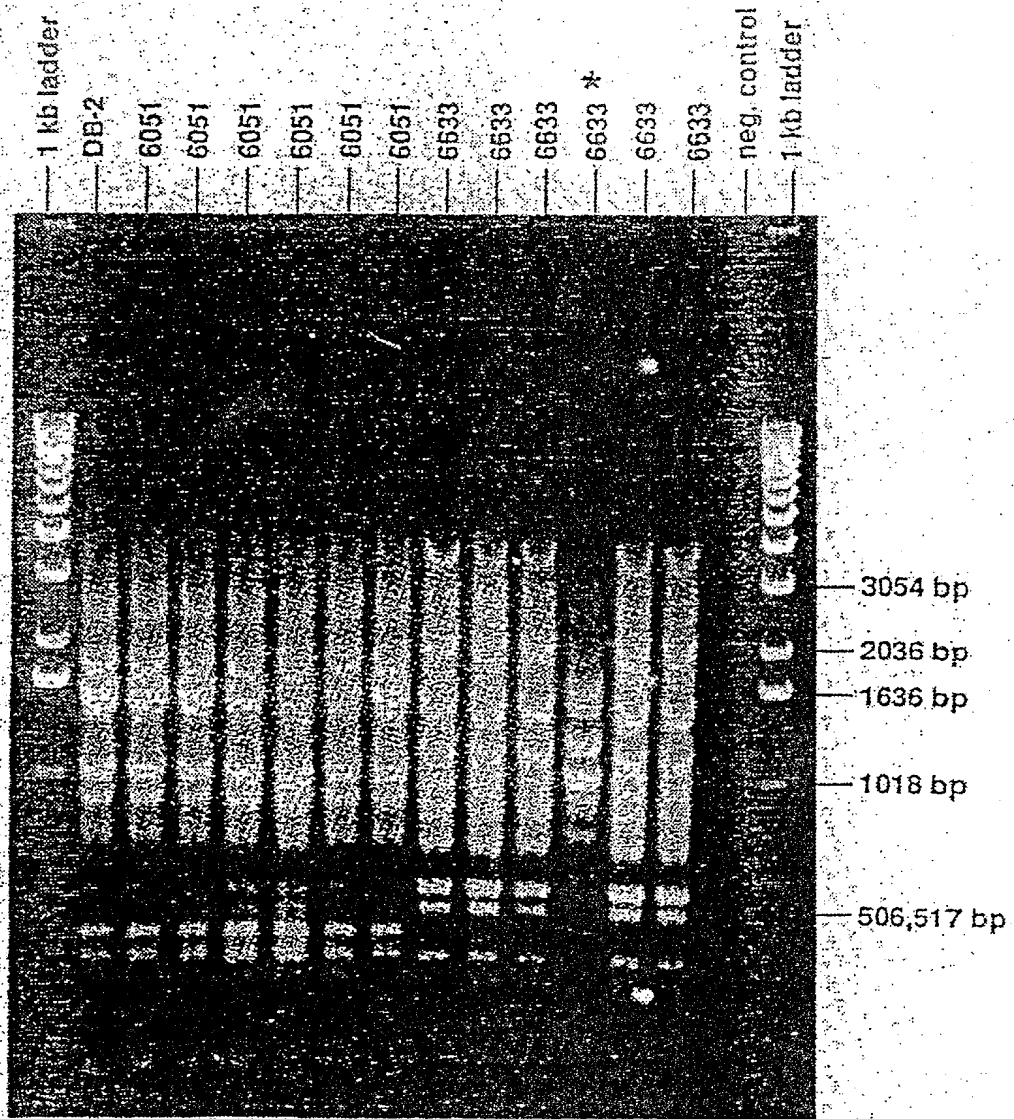


Figure 13

18/19

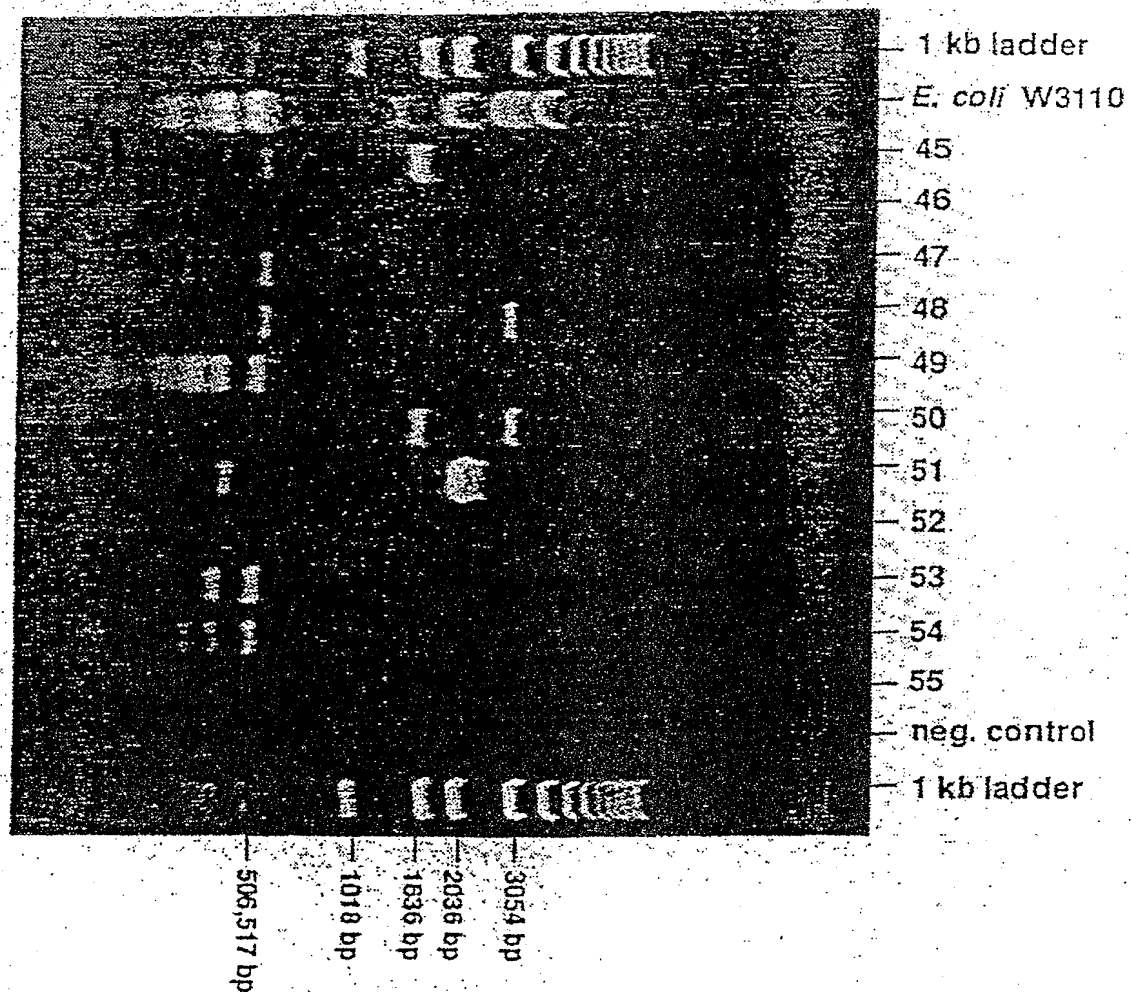


Figure 14

19/19

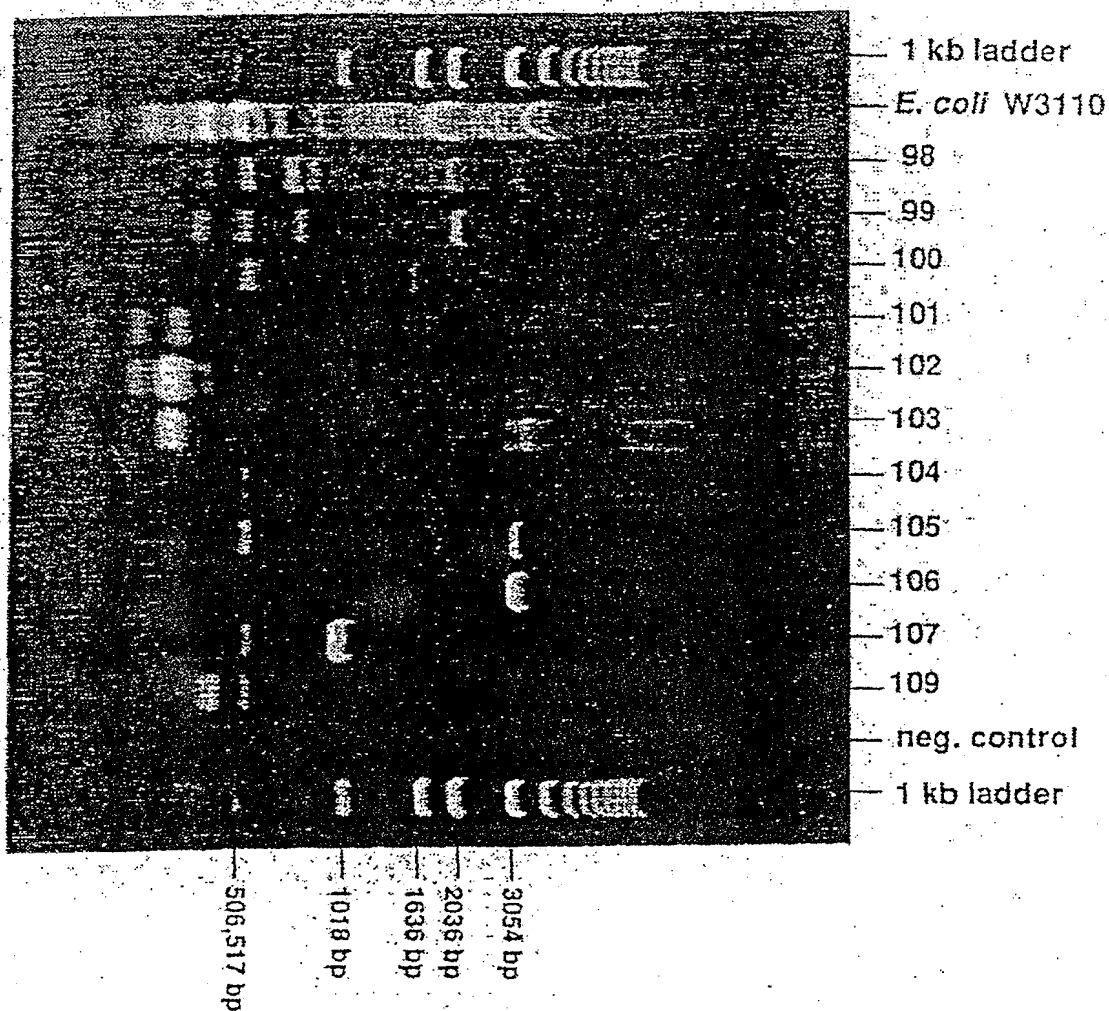


Figure 15



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US92/09230

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12P 19/34; C12Q 1/68; C07H 21/04; C12M 1/38

US CL : 435/91, 6, 290; 536/24.33

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91, 6, 290; 536/24.33; 935/77, 78, 79, 87

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS-ONLINE, BIOSIS, MEDLINE, APS, GENBANK, UEMBL

search terms: repetitive sequences, REP, ERIC, fingerprint, DNA, bacteria

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                   | Relevant to claim No.                                |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|
| X,P       | NUCLEIC ACIDS RESEARCH, Volume 19, No. 24, issued December 1991, J. Versalovic et al., "Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes", pages 6823-6831, see entire document.                                                                                                                                        | 1-60,69-72,86-88                                     |
| X,P<br>Y  | APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Volume 58, No. 7, issued July 1992, F. J. de Bruijn, "Use of Repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of <u>Rhizobium meliloti</u> isolates and other soil bacteria", pages 2180-2187, see entire document. | 1,3,4,15-21,<br>30-60,86<br>2,5-14,22-29,69-81,87-88 |
| X         | JOURNAL OF BACTERIOLOGY, Volume 174, No. 14, issued July 1992, G. P. Dimri et al., "Physical Mapping of Repetitive Extragenic Palindromic Sequences in <u>Escherichia coli</u> and phylogenetic distribution among <u>Escherichia coli</u> strains and other enteric bacteria", pages 4583-4593, see entire document.                                                                | 61                                                   |
| Y,E       | US, A, 5,176,203 (Larzul) 05 January 1993, see entire document.                                                                                                                                                                                                                                                                                                                      | 62-68,82-85                                          |

☒

Further documents are listed in the continuation of Box C.

☐

See patent family annex.

|                                        |                                                                                                                                                                     |   |                                                                                                                                                                                                                                              |
|----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Special categories of cited documents: |                                                                                                                                                                     | T | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principles or theory underlying the invention                                             |
| *A*                                    | document defining the general state of the art which is not considered to be part of particular relevance                                                           | X | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| *E*                                    | earlier document published on or after the international filing date                                                                                                | Y | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L*                                    | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | X | document member of the same patent family                                                                                                                                                                                                    |
| *O*                                    | document referring to an oral disclosure, use, exhibition or other means                                                                                            |   |                                                                                                                                                                                                                                              |
| *P*                                    | document published prior to the international filing date but later than the priority date claimed                                                                  |   |                                                                                                                                                                                                                                              |

Date of the actual completion of the international search

16 JANUARY 1993

Date of mailing of the international search report

16 FEB 1993

Name and mailing address of the ISA/  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

LISA BENNETT

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US92/09230

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                     | Relevant to claim No.                                           |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Y         | US, A, 5,038,852 (Johnson et al.) 13 August 1991, see entire document.                                                                                                                                                                                                                                                                 | 62-68, 82-85                                                    |
| X<br>Y    | NUCLEIC ACIDS RESEARCH, Volume 19, No. 4, issued January 1991, J. Welsh et al., "Genomic fingerprints produced by PCR with consensus tRNA gene primers", pages 861-866, see entire document.                                                                                                                                           | 1,3,4,15-16,18-20,<br>30-60<br>2,5-14, 17,21-29,69-<br>81,86-88 |
| Y         | GENOMICS, Volume 6, No. 3 issued March 1990, S. A. Ledbetter et al., "Rapid Isolation of DNA probes with Specific chromosome regions by interspersed repetitive sequence polymerase chain reaction", pages 475-481, see entire document.                                                                                               | 1-60,69-81,86-88                                                |
| Y         | Cell, Volume 73, issued July 1984, M.J. Stern et al., "Repetitive extragenic palindromic sequences: a major component of the bacterial genome", pages 1015-1026, see especially Figure 1.                                                                                                                                              | 1-5,8,12-60, 69-<br>75,79,86-88                                 |
| Y         | MOLECULAR MICROBIOLOGY, Volume 5, No. 4, issued 1991, C.S. Hulton et al., "ERIC sequences: a novel family of repetitive elements in the genomes of <u>Escherichia coli</u> , <u>Salmonella typhimurium</u> and other enterobacteria", pages 825-834, see especially page 825, Figure 1, page 828, col. 2, Figure 4, and pages 830-833. | 1-6,8-9,12-60,69-76,79-<br>80, 86-88                            |
| Y         | JOURNAL OF BACTERIOLOGY, Volume 167, No. 3, issued September 1986, F. F. Correia et al., "A 26-base-pair repetitive sequence specific for <u>Neisseria gonorrhoeae</u> and <u>Neisseria meningitidis</u> genomic DNA, pages 1009-1015, especially the abstract and figure 4.                                                           | 1-4,7,10,12-60,69-<br>74,77,81,86,88                            |
| Y         | JOURNAL OF BACTERIOLOGY, Volume 173, No. 6, issued March 1991, E. Lennon et al., "A highly conserved repeated chromosomal sequence in the radioresistant bacterium <u>Deinococcus radiodurans</u> SARK", pages 2137-2140, especially Figure 3.                                                                                         | 1-4,11-60, 69-74,<br>78,86,88                                   |